Exploration of novel two dimensional materials by using first principle calculation and materials informatics approach

Itsuki Miyazato

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

of

Hokkaido University.

Graduate School of Engineering

Hokkaido University

February 15, 2019

Contents

List of Figures

List of Tables

Introduction

Materials Informatics (MI) gains recently remarkable attention from both academic and industrial field with the goal of efficient material research. A traditional approach for material development is carried out in research laboratory with hidden failure experiment and its feedback by individual specialized researchers. This approach normally requires time and cost to develop practical materials due to its try-and-error process, but at the same time, this costly error process can be a great information for the next success. The material design now is advanced with a development of computation equipment and methodology; some specified systems can be simulated by computation with sufficient accuracy, such as fluid dynamics and chemical dynamics. In the field of materials design, density functional theory(DFT) is one of the revolutionary methods for novel materials modeling since it can provide various kind of information from its calculations, for example, band structure for electronic device, bulk modulus for structure materials, reaction path evaluation for catalysis, etc. In particular, material predictions have been achieved by DFT calculations against nano materials, such as nanoclusters, nanotubes and two dimensional materials. And some of them are successfully synthesized and applied into functional materials, especially, two dimensional materials are remarkable nanomaterials because of its unique properties with wide range application: from mechanical to medical usage. However, simulation has still disadvantage of its accuracy and implementation for real material design, therefore, it cannot be the most efficient method for material design. Consequently, simulations still rely on researchers experience in the process of designing the new model materials, even though this process can be replaced costly experimental process. MI is one covering

tool for this empirical side in material design, predicting new materials from exist data by applying data science. This method has great potential to discover new materials since the all prediction process is conducted by stiff computer algorithm, consequently, all data is treated equally unlike researchers inspiration. However, MI has no established concrete method because of its novelty. In this study, both computational(DFT) and MI method is implemented to discover novel two dimensional materials.

In the 1st chapter, the overview of MI is explained: from the history to its concept and outlook to the future research. The current progress in MI among both academic and industries are explained.

In the 2nd chapter, two dimensional materials, target materials of this research is introduced in the view of its history and remarkable properties.

In the 3rd chapter, the fundamentals of first principle calculation is introduced. The fundamental equations and theories are explained.

In the 4th chapter, the implementation of calculation code is explained. In particular, GPAW code, a DFT calculation code applied in the studies, are explained from the point of view to feature functions, installation and high throughput calculation.

In the 5th chapter, the research result achieved by first principle calculation is introduced. Hexagonal boron nitride (h-BN) is commonly known substance with graphite-like layered materials. And two-dimensional boron phosphide (h-BP) is predicted its existence and semiconductive property by first principle calculation. Group XV atoms are common in these h-BN and h-BP two-dimensional materials, hence, boron with Group XV atoms(M: As, Sb and Bi) are investigated and evaluated by first principle calculations due to designing novel materials. As a result, those single layer boron based novel two-dimensional materials turned out to be energetically stable, and their electronic property suggests a potential application for semiconductor and catalyst. Band gap and binding energy in these single layer boron based two-dimensional materials turned out to be proportionate to the Allen electronegativity. Additionally, the layered structures are investigated, and zigzag formation is proposed in B-Sb and B-Bi. And electronic structure turned out to

differ from single layer materials.

In the 6th chapter, the fundamentals of machine learning is introduced for the explanation for ML method within this research. In particular, the variation of machine learning method is explained in the views of ML implumentation, and Naive Bayes, implemented algorithm in the study, is introduced.

In the 7th chapter, the implementation of machine learning is explained. In particular, scikit-learn, a machine learning module for machine learning in Python language program is introduced for MI implementation.

In the 8th chapter, the combined research result with first principle calculations and MI is introduced. novel two-dimensional magnets are explored by materials informatics approach and investigated by first principle calculations. 216 computational two-dimensional material data within an open database is collected and explored by machine learning. Gaussian naive Bayes algorithm is applied in order to predict MoS_2 shaped(AB_2) and graphene shaped(AB) two-dimensional magnets with high magnetic moment. As a result, novel 254 AB and AB_2 two dimensional materials. are explored as candidates with high magnetic moments. By the evaluation of first principle calculations, 7 candidates are found to be energetically stable and have high magnetic moment: $MnPd_2$, FeS, CrSe, CrS, MnTe, MnSe, and MnS.

In the 9th chapter, the research achievement are summarized. And perspective of MI is discussed the feedback of this research. The both advantage and disadvantage of theoretical data is explained in the consideration of doctoral studies. And future perspective of MI is described by the consideration of experimental data implementation.

Chapter 1

Materials Informatics

1.1 History

Materials Informatics (MI) is a new stream of materials design technique originated from the “Materials Genome Initiatives” project in the United States. MI is the combined field of “Materials Science” and “Informatics” , which aims to reduce time and cost for material development and practical application by applying data science into material development. In traditional approach, new materials are designed, synthesized, and evaluated by researchers’ inspiration or experience. Some innovative materials, such as “Polytetrafluoroethylene(known as Teflon^{TM})”, “Neodymium magnet(strongest permanent magnet; discovered by Masato Sagawa)”,“Conductive polymers(Nobel Prize in Chemistry by Hideki Shirakawa, Alan MacDiarmid and Alan Heeger.), have been discovered by chance. This discovery process stands on assumption of human, which generally requires ”try-and-process“. For example, researchers are firstly design materials with his/her assumption imaginary, then validate this assumption by experiment: synthesizing and evaluating materials. Then, the feedback is obtained from the experimental result and rebuild next strategy. This process continues until the desired materials developed and evaluated, therefore, this approach sometimes cost time and money. In recent years, performance of computers has dramatically improved, as well as the computer simulation techniques,such as first principle calculations, have been developed. Consequently, some costly process is alternated by computer simulation in materials design, how-

ever, the ”try-and-error“ framework is still stand.

The raise of Informatics suggests one promising sight in Materials Science. James Nicholas Gray(1944–2007), an American computer scientist, proposed that this age is shifting to 4th paradigm in science.

1, Empirical approach (thousand years ago) Observation of natural phenomenon and describing its mechanism. Example: Discovery of heliocentrism, stellar parallax, Mendelian inheritance, etc.

2, Theoretical approach (last few hundred years) Building models, or generalizing observed phenomenon by mathematical equation. Example: Newtonian mechanics, quantum physics theory, electromagnetics, logic gate, etc.

3, Computational approach (last few decades) Simulating complex phenomenon by computation. Example: molecular modeling, fluid dynamics simulation, weather forecast, electrical circuit modeling, etc.

4, Data exploration approach (nowadays) Gathering data from theory, experiment and simulation.This research field is generally called “XX + informatics” Example: Bioinformatics, Materials informatics, Chemoinformatics, Astroinformatics, etc.

As his prediction, in these days, the interdisciplinary fields, which is commonly expressed as “XX + informatics” is raised in various fields. The field of Material Science is not exception: the first movement of Materials Informatics has begun in the United States. Combinatorial approach is firstly considered as the potential method to shorten material development period, such as electronics, catalysis and coatings. National Institute of Standard and Technology(NIST), a non-regulatory agency of the United States of Commerce, built NIST Combinatorial Materials Center to develop organic material synthesis, high-throughput screening and methodology of data accumulation. On the other hand, the National Science Foundation(NSF) launched “Combinatorial Science and Materials

Informatics Collaboratory(CoSMIC) to develop high-throughput experiment and data analysis methodology by using mathematical and statistical approach. For the Integration of obtained knowledge from those project, ”Materials Genome Initiative(MGI)“ was established on 2011 as the national project of the United States under Barak Obamas's administration, in order to find a breakthrough method for material design by applying data science approach. The one notable achievement of MGI is discovery of solid state electrolyte for Li-ion battery without any experimental try-and-error process: The research group of MIT and Samsung electronics discovered the novel solid state electrolyte by using computer simulations and data science approach[3]. MGI researchers said to be deeply involved this research. After the establishment of MGI, countries all over the world begin to establish own project related to Materials Informatics:

- **European Union : NOMAD**

  NOMAD(Novel Materials Discovery)[4] is representative laboratory to discover materials by big data analysis and computational simulation. This laboratory maintains database of input and output files of computational codes to share among countries and research institutes, especially focused on functional materials such as solar cells, structural materials and candidates of batteries.

  Switzerland has also established AiiDA[5] with the aim to collect electronic structures calculated by first principle calculations.

- **China : Chinese MGI**

  China has also undertaken for Chinese ”MGI“ project under the Chinese Academy of Sciences[6]. In the educational side, Shanghai University established ”Materials Genome Institute“ to train for the expert of Materials Informatics[6].

- **Japan : MI^{2}I**

  Japan has launched ”Materials research by Information Integration Initiative(MI^{2}I)“

in National Institute of Material Science(NIMS) under the project of Japan Science and Technology Agency(JST). MI^2I is mainly focused on developing storage battery materials, magnetic materials, thermal management materials and thermoelectric materials. MI^2I is aims to establish under the goals below[7]:

- Creating a system of industryacademiagovernment collaboration over a wide range of areas, from materials science to information and mathematical sciences
- Promoting the participation of companies from various fields in the creation of an open innovation hub for work on data-driven (information-integrated) materials research.

The achievement by Materials Informatics is also raised in recent years:

In Japanese research, several research group achieved to find out new materials by applying first principle calculation and Materials Informatics approach. The research group of SHARP Corporation and Kyoto University discovered new lithium iron phosphate cathode in 2014[8] The research group of Tokyo Institute of Technology succeeded to develop novel red light emitting diode without rare earth[9] The experimental approach is also ongoing by Materials Informatics approach, such as high-temperature ferroelectric perovskites[10],

Industries also begin to work on into Materials Informatics to develop their products with efficiency. BASF(World's largest chemical company) has began research collaboration with Citrine Informatics to use artificial intelligence (AI) to accelerate the development of new environmental catalyst technologies[11]. Toyota research institute in the U.S. has invested 35 million dollars for 4 years research to help accelerate the design and discovery of advanced materials by using AI[12]. In Japanese companies, Fujitsu, NEC and Hitachi has started to apply Materials Informatics approach in order to develop batteries and heating materials[13]. Fujitsu has established special team for Materials Informatics in device and material laboratory. By combing Fujitsu's technology of supercomputer, a tons of material data has been

collected by first principle calculation for data-mining.

1.2 The material data handling

Materials Informatics fundamentally needs material data in advance(Figure 6.2). This data includes both experimental and theoretical(simulation). One of the most important data in database is negative data. Researchers tend to pay attention to materials with remarkable properties because it will be the clue of novel material discovery. However, in Materials Informatics approach, negative data will function to identify ”less promising“ candidates by analyzing data structure of whole material database. The existing materials data is preprocessed due to adaptation of machine learning.

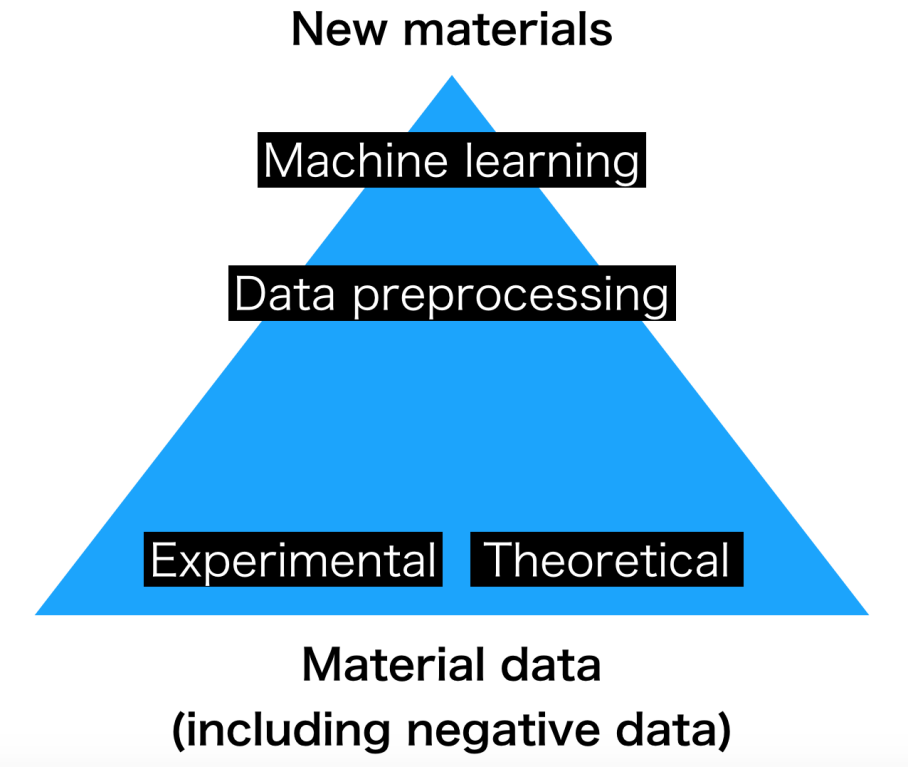

Figure 1.1: The image of Materials Informatics.

Data acquisition method is first and important point for Materials Informatics since the quality of data is strong factor for material prediction. In the data science, ideal data is identified in the following 3 aspects.

- **Quality of data**

Fixed quality is needed. If the most of data is fake, or data quality is sparse, nature of data cannot be captured.

- **Homogeneity of data**

  Dispersed data is preferable to capture the nature of data, in other words, negative study data could be helpful.

- **Quantity of data**

  Large quantity of data can hide negative aspect(such as internal error) of data statistically.

In the material data, ”experiment“ and ”simulation“ data is exist. These two data has each character in the sight of data science, so the data should handle in ”the right place“.

Simulations

Simulation can provide ideal data for informatics in the sight of its homogeneity, flexibility and reproductivity.

The most common method for data acquisition is first principle calculation in these days. First principle calculation provides various kind of information in one calculated results, such as structure parameters, electronic properties, bulk modulus, reaction paths etc. However, simulation provides limited information by its nature. For example, DFT calculation result fundamentally cannot include the thermal effect, and accuracy of result is dependent on constituent atoms, or its calculation parameters. Additionally, DFT calculation models contain no aspects for material design, for example, unrealistic structure for synthesis, lack of aspects in material functions(e.g. thermal effect in catalysis) etc.

Experimental

The experimental data is the most realistic data for material design., which can provide experimental condition information simulation cannot provide, such as reaction pressure, temperature. This aspect has already reported in the oxidation coupling of methane(OCM) catalyst[14]. However, existing experiment data has disadvantage for implementation to Materials Informatics in the views of following 3

aspects:

- **Quality**

  The data quality of experiment relays on actual operator of experiment(sample preparation, data collection, etc). His/her skills and goal for the experiment directly reflects on the quality of data.

- **Homogeneity**

  The reported experimental data is normally concentrated in the field of "good" result. Experiment cost time and money so that experiment is not usually carried out to obtain "negative" data intentionally, and negative data is seldom reported on the paper.

- **Dependency of analyzing instrument**

  Accuracy of analyzing instrument is different from each laboratory(calibration, analyzing method, etc) and individual instrument.

Therefore, reported existing data, such as accumulated data by text mining from papers, are carefully handled for the data processing process. The one solution for this disadvantage can be developing high-throughput experiment. This experiment approach provides experimental data with homogeneous data with the same analyzing instruments rapidly. The rapidness also encourages to collect negative data intentionally, which is often avoided or ignored in conventional experiment. Hence, the development of high-throughput experiment technique is highly required for the advance of Material Informatics.

The collected data is properly organized in various usage aspects, such as unified units, standardized formats, guaranteed accuracy, etc. The important concept of building database can be "Ontology". Ontology is one field of information science, which defines the categories, properties, relations and relations in certain data. The one successful example of ontology is gene ontology(GO)[15] in Bioinformatics. GO realized the cross-comparison and combining of different types of gene information obtained from different institution since containing data is well organized and described with each relationship by ontology. By this example, building

ontology on material data is one of the most important key to accelerate material discovery by Materials Informatics in the views of collecting ”useful“ data Materials data ontology is recently being advocated in the sight of catalyst design[16].

1.3 Spectroscopy analysis in data science

The another problem of handling experimental data is mass analysis techniques. High-throughput experiment can provide experimental data quickly, including spectroscopy data. However, spectroscopy analysis still relies on experimental operator because of its complexity. One potential solution is applying informatics approach on spectroscopy analysis. The development of high-throughput analyzing is now ongoing issue by the implementation of informatics into analytical science, such as X-ray near edge spectrum(XANES) in XAS[17, 18, 19, 20], Raman spectroscopy[21, 22], and the research achievements are now frequently reported in these days. The database construction is also ongoing for informatics approach, such as X-ray absorption spectroscopy(XAS)[23].

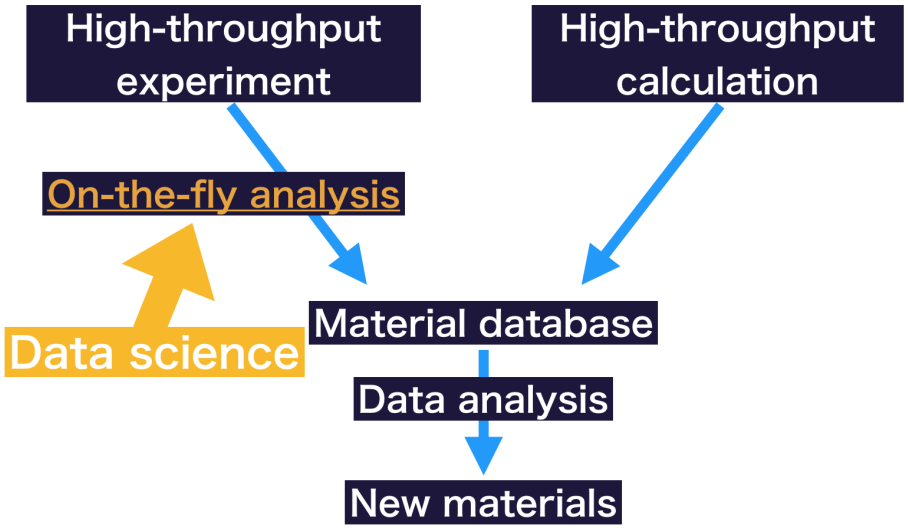

Figure 1.2: The role of data-driven spectroscopy analysis in Materials Informatics.

Chapter 2

Two dimensional materials

Two dimensional materials are commonly popular crystalline materials with single layer of atoms. The most popular two dimensional materials are graphene, one of the carbon allotrope consisting plane single layer of carbon atoms with honeycomb structure. Since the discovery of graphene in unique properties[24, 25], its unique properties has been focused to develop novel materials, such as electronics, semiconductor, battery and energy conversion[26]. On the other hand, another types of two dimensional materials are predicted to be theoretically exist by first principle calculations[27]. In this section, two dimensional materials, main target of this research, are introduced.

2.1 Variation of two dimensional materials and their properties

Since the discovery of graphene, various types of two dimensional materials are synthesized experimentally, and tons of potential two dimensional materials are proposed by first principle calculations.

2.1.1 Synthesized compounds

- **Graphene**

  Graphene was discovered by Andrei Konstantinovich Geim, a Russian physicist(1958-), in 2004. Regarding to those research achievement, Geim won Nobel Prize in Physics with Konstantin Novoselov, as “for ground-

breaking experiments regarding the two-dimensional material graphene"[28]. The word itself, "Graphene", is formerly existed as the combination word of "Graphite" and the suffix -ene. It is firstly named by Hanns–Peter Boehm[29], who observed single–layer carbon foil in 1962[30]. Geim is defined "Graphene is a single atomic plane of graphite, whichand this is essentialis sufficiently isolated from its environment to be considered free-standing"[31].

Several methods are available for the synthesis of graphene. Graphite is commonly known to have structure consisting multi–layered graphene with van der Waals force between layers. Hence, Geim and Novoselov firstly used adhesive tape to peel from bulk graphite to graphene flakes and transferred them to a silicon wafer[28]. Structure of graphene is well known as the plane honeycomb–like single sheet consisting with carbon atoms. The distance of atoms are 1.42 Å[32]. First principle calculation suggest that a graphene sheet needs more than 6000 carbon atoms to be thermodynamically stable[33]. This instability in two dimensional crystal is considered for the reason of "rippling" of the flat sheet in graphene[34, 35].

- **Hexagonal boron nitride(h-BN)**

  Hexagonal boron nitride(h-BN), polymorph of boron nitride, is commonly known because of its graphite-like layered structure. Weak van der Waals force is exist in its interlayer, as similar as graphene. Hence, h-BN has also known as the name "white graphene". In the application aspects, h-BN is firstly commercialized in cosmetics around 1940 in Japan due to its white color. Since h-BN is chemically and thermally high stable, its ceramics are widely applied to the materials for high-temperature environment. The another implementation is as hybrid materials, such as alloys, resins, ceramics, rubbers, by the usage of its self-lubricating properties. These hybrid materials are applied to make bearings or be used in steel industries. Thus, h-BN is can be defined as the most used two dimensional structured materials. The hybrid material of graphene and h-BN is called "Borocarbonitrides", which

have B-C, B-N, C-N, and C-C bonds in its honeycomb ring[36, 37]. Borocarbonitrides varies its band gap(1.0-3.9 eV[38]) by tuning the content of the carbon and boron nitride domains, considering promising material in electronic devices.

- **Transition metal dichalcogenides(TMDCs)**

  Transition metal dichalcogenides(TMDCs) are also commonly known two dimensional crystal compounds, such as MoS_2[39], $MoSe_2$[40], $MoTe_2$[41, 42], WS_2[43], and WSe_2[44]. MoS_2 has direct band gap and application to transistors and in optics as emitters and detectors are proposed[45, 46, 47, 48].

- **Germane**

  Germane is a graphene-like single layer material composed of germanium one hydrogen bonded in the z-direction for each atom[49, 50]. Germane was firstly synthesized by molecular beam epitaxy on (111) gold surface, and structure was observed with scanning tunneling microscopy[51]. The property was evaluated by first principle calculation and it revealed to be potential material candidate for field-effect transistors[52]

- **Bithmuthene**

  Bithmuthene is a two dimensional materials of Bismuth(Bi), synthesized on SiC(0001) surface[53] and predicted to be a topological insulator by first principle calculation.

- **Silicene**

  Silicene is a two dimensional materials of Silicon(Si) with graphene-like honeycomb structure, however, it have periodically buckled unlike graphene.

2.1.2 Theoretically predicted materials

First principle calculation is commonly used simulation method to understand the physical properties of nano materials. This simulation method provides various type of physical parameters within its simulated result(refer Chapter 3 and Chapter 4), hence, the exploration of potential nano materials are achieved, including two

dimensional materials. It is reported that potential 700 two dimensional materials have been predicted to be energetically stable by first principle calculations[27].

- **Si_2BN**

  Si_2BN is a theoretically predicted two dimensional materials by first principle calculation. This material is considered to have metallic properties with only sp^2 bonds[54].

- **Transition metal trichalcogenides (TMTs)**

  TMTs are promising two dimensional material candidates in photoelectronic properties predicted by first principle calculations, with the composition of MX_3(M=transition metal, X=chalcogenide)[55, 56]. TMTs is considered as the potential candidate for photocatalytic water splitting[57, 58], optoelectronic devices[57, 58, 59, 60, 61], Thermoelectric (TE) materials[62] and solar energy conversion[57] energy conversion.

Chapter 3

Fundamentals of density functional theory

Density functional theory (DFT) is one of the most popular method, and commonly used in computational physics. DFT is powerful tool to describe the properties of condensed matter systems due to its availability to various kind of materials, from bulk to nano-scaled matter. DFT method stands on first principle theory based on Schrödinger equation. Schrödinger equation is expressed as wave function, describing quantum state in certain condition. However, Schrödinger equation itself is impossible to solve because of N-body problem. In the natural condition, many-body system is common so that approximation technique had been invented to describe real quantum system. DFT was developed as the solution to overcome this problem in order to reduce the $3N$ degree of freedom if the N body system by assuming its electron density in three spacial coordinates. In this section, the basic idea of DFT is given to introduce how the properties of designed materials are obtained from calculated result.

3.1 Born-Oppenheimer Approximation

The first approach to solve many-body problem in quantum physics was achieved by Born-Oppenheimer Approximation by assuming that nuclei is instantaneous. This approximation is based on the idea of justification that nuclear motion ($\approx 10^3$ m/s) is negligible by the comparison of electronic motion ($\approx 10^6$ m/s). This approximation

implies that the solution of electron wave function is expressed as fixed Schrödinger equation for Hamiltonian:

$$\hat{H} = O[\rho] \tag{3.1}$$

3.2 HohenbergKohn theorems

1st theorem

The ground state density of many electron system ($\rho(\overrightarrow{r})$) is correspondent to the external potential(V_{ext}) The ground state expectation value ($\hat{O}$) can be described as a unique function of the exact ground state electron density(equation(3.2)).

$$< \psi|\hat{O}|\psi >= O[\rho] \tag{3.2}$$

2nd theorem

The ground state total energy ($E_{Vext}[\rho] \equiv H[\rho]$) can be described as equation(3.3) if ($\hat{O}$) is Hamiltonian.

$$E_{Vext}[\rho] =< \psi|\hat{T} + \hat{V}|\psi > + < \psi|\hat{V}_{ext}|\psi > \tag{3.3}$$

$$= F_{HK}[\rho] + \int \rho(\overrightarrow{r}) V_{ext}(\overrightarrow{r}) d\overrightarrow{r} \tag{3.4}$$

- $E_{Vext}[\rho]$: The ground state total energy(=Minimal value for the ground state)
- $F_{HK}[\rho]$: Hohenberg-Kohn density functional

These theorem implies that ground state density is correspondent to the external potential. As considering 1st theorem, the electron density contains as much information as the wave function. Consequently, 2nd theorem implies that ground state density minimizes $E_{Vext}[\rho]$, which can be obtained by applying Rayleigh-Ritz method. The equation is the same for all many-electron system since the value $F_{HK}[\rho]$ contains no terms of the nuclei and nuclear positions However, these theorem didn't provide the way to calculate the ground state density because they didn't give the value of $F_{HK}[\rho]$. Therefore, the Kohn-Sham theory is introduced to solve this $F_{HK}[\rho]$, which is the fundamental key to develop DFT.

3.3 Kohn-Sham equations

Kohn and Sham developed "Kohn-Sham equation" in 1965[63] to compute ground state, by providing the solution of $F_{HK}[\rho]$. In this theory, $F_{HK}[\rho]$ was rewrote as the following equation (3.5)

$$F_{HK}[\rho] = T_0[\rho] + V_H[\rho] + V_{xc}[\rho] \tag{3.5}$$

- $T_0[\rho]$: the functional of kinetic energy for non-interaction electrons.
- $V_H[\rho]$: Hartree contribution.
- $V_{xc}[\rho]$: Exchange correlation functional

This theorem, however, considers no electron-electron interaction, $V_H[\rho]$ approximates in electron interaction since it describes with the field obtained by averaging at the position of existing electrons. $V_{xc}[\rho]$ describes the effect originated from electron exchange interaction and electron correlation. Hence, $E_{Vext}[\rho]$ can be expressed as

$$E_{Vext}[\rho] = T_0[\rho] + V_H[\rho] + V_{xc}[\rho] + V_{ext}[\rho] \tag{3.6}$$

This equation can be interpreted with the Kohn-Sham Hamiltonian as follows:

$$\hat{H}_{ks} = \hat{T}_0 + \hat{V}_H + \hat{V}_{xc} + \hat{V}_{ext} \tag{3.7}$$

$$= -\frac{\hbar}{2m_e}\nabla_i^2 + \frac{e^2}{4\pi\varepsilon_0}\int \frac{\rho(\hat{\overrightarrow{r'}})}{|\overrightarrow{r} - \hat{\overrightarrow{r'}}|} d\hat{\overrightarrow{r'}} + + \hat{V}_{xc} + \hat{V}_{ext} \tag{3.8}$$

The exchange correlation operator $\hat{V}_{ext}$ can be expressed as the functional derivative:

$$\hat{V}_{xc} = \frac{\partial V_{xc}[\rho]}{\partial \rho} \tag{3.9}$$

Therefore, Kohn-Sham theorem can be stated "The exact ground state density $\rho(\overrightarrow{r})$ of N-electron system can approximate as

$$\rho(\overrightarrow{r}) = \sum_{i=1}^{N} \psi_i^*(\overrightarrow{r})\psi_i(\overrightarrow{r}) \tag{3.10}$$

where the single particle wave function $\psi_i(\overrightarrow{r})$ are the N lowest energy solutions of the Kohn-Sham equation."

$$\hat{H}_{ks}\psi_i = \varepsilon_i\psi_i \tag{3.11}$$

3.4 Exchange-correlation functional

As explained in the previous section, exchange correlation functional, which express the effect originated from electron exchange interaction and electron correlation, need to define to solve the Kohn-Sham equation. However, exact expression of this effect is unavailable so the approximation approach is introduced. The oldest approximation is LDA (Local Density Approximation), which is defined as the following equation.

$$V_{xc}^{LDA}[\rho] = \int \rho(\overrightarrow{r})\varepsilon_{xc}(\rho(\overrightarrow{r}))d\overrightarrow{r} \tag{3.12}$$

where $)\varepsilon_{xc}(\rho(\overrightarrow{r}))$ is the exchange correlation function. This approximation stands on the idea that the each point of exchange correlation energy is approximated locally by exchange correlation energy of homogeneous electron gas with the same electron density. LDA provides good results for geometrical quantities in atoms and simple molecule systems, such as bond length, electron densities, vibrational frequencies and energy differences (ex: ionization potentials)[64]. However, LDA overestimates the ground state energy in open shell atom system.

GGA (Generalized Gradient Approximation) approximation[65, 66, 67] is developed to overcome those defect of LDA approximation due to introducing the idea of density gradient in its approximation.

$$V_{xc}^{GGA}[\rho] = \int \rho(\overrightarrow{r})\varepsilon_{xc}(\rho(\overrightarrow{r}),|\nabla\rho(\overrightarrow{r}))d\overrightarrow{r} \tag{3.13}$$

The function $\varepsilon_{xc}(\rho(\overrightarrow{r}),|\nabla\rho(\overrightarrow{r}))$ is variously defined and many forms are pro-

posed. Therefore, derivations has been developed to improve calculation accuracy, such as PW91[67] and B3LYP[68]. In this study, GGA approximation of Perdew Burke Ernzerhof (PBE) exchange correlation function is applied [69], which is commonly used exchange correlation functional due to its accuracy and

3.5 Linear combination of atomic orbitals

Linear combination of atomic orbitals(LCAO) is a calculation approximation method implemented in first principle calculation. This method approximates certain monocular orbital as atomic orbitals of constituent atoms by linear conbination. In mathematical form, therefore, can be expressed as the following equation where n atomic orbitals are

$$\phi_i = c_{1i}\chi_1 + c_{2i}\chi_2 + c_{3i}\chi_3 + \cdots + c_{ni}\chi_n \tag{3.14}$$

$$= \sum_r c_{ri}\chi_r \tag{3.15}$$

where ϕ_i, χ_r

This LCAO method is implemented in SIESTA (Spanish Initiative for Electronic Simulations with Thousands of Atoms) code[70], as well as in GPAW code, which has the most cheapest computational cost among implemented wave functions in GPAW code. Therefore, all initial calculation(finding lattice constant, structure relaxation) is performed by LCAO method to obtain “rough” result for initial evaluation in the sight of reduction for calculation time.

3.6 Density of states

Density of states(DOS) is a effective diagram to evaluate electronic structure from the first principle calculation result. The density of states is defined by the following equation:

$$\rho(\varepsilon) = \sum_n < \psi_n|\psi_n| > \delta(\varepsilon - \varepsilon_n) \tag{3.16}$$

where ε is number of electronic states with energy.

This equation can be rewritten as follows by acting with the unit operator:

$$\rho(\varepsilon, r) = \int dr \sum_n < \psi_n | r >< r | \psi_n > \delta(\varepsilon - \varepsilon_n) \tag{3.17}$$

This $\rho(\varepsilon, r)$ is called the local density of states(LDOS). Additionally, it is useful to obtain the projected density of states(PDOS) within well known basis function such as atomic prbitals when the electronic structure is evaluated. PDOS can be defined by the following equation:

$$\rho(\varepsilon) = \sum_n < \psi_n | \phi_i^a >< \phi_i^a | \psi_n > \delta(\varepsilon - \varepsilon_n) \tag{3.18}$$

3.7 Bader charge analysis

The ground state electron density contains the information of the location of electrons. The Bader charge analysis[71, 72, 73] is commonly applied method. The electron density is divided by zero flux surfaces.

$$S = \{ r \in \mathbb{R}^3 | \nabla n(r) / u(r) = 0 \} \tag{3.19}$$

3.8 Adsorption energy and interlayer binding energy

The adsorption energy(E_{ads}) is good indicator to consider the interaction between surface and adsorbate. The adsorption energy can be calculated by comparison of total energy difference:

$$E_{ads} = E_{adsorbate+slab} - (E_{adsobate} + E_{slab}) \tag{3.20}$$

where $E_{adsorbate+slab}$, $E_{adsorbate}$ and E_{slab} indicate total energy of slab and adsorbate, adsorbate alone, and slab alone, respectively. Additionally, the interlayer binding energy E_{int} between facing layers are calculated by the following equation:

$$E_{int} = E_{total} - N \times E_{1layer} \tag{3.21}$$

where E_{total}, N and E_{1layer} express calculated total free energy of layered material, number of layer in material (≥ 2), and total free energy of single layer material.

Chapter 4

Implementation of DFT calculation code : GPAW

DFT calculation code is developed and distributed by various organizaitons. The most popular commercial DFT calculation code is Vienna Ab initio Simulation Package(VASP)[74, 75], CASTEP and Gaussian[76]. For the non-commercial code, Quantum Espresso[77, 78] and SIESTA[70] are popular and widely used.
However, in this study, GPAW[79, 80] code is implumented since it is highly integrated with Python language[81, 82].This feature allows to automate high-troughput calculation process since Python integrates many data processing modules(the advantages of Python into data science is described in Chapter7) by building Python script. Subsequently, GPAW is provided by GNU Public License version 3, and able to install on Linux operating system. Therefore, the calculation environment can be built without any software license fee.

4.1 GPAW

The following analysis is performed by GPAW code in order to evaluate the electronic structure of designed two dimensional materials.

- Densiy of states (DOS) : Evaluation of electronic structure of two dimensional materials.
- Bader charge analysis : Considering the charge transfer between target atoms. Bader charge output is calculated by FORTRAN code provided from Henkel-

man Group[83] from the Gaussian CUBE format, which can be obtained from GPAW calculation.

- Adsorption energy : Considering the interlayer interaction in multi-layered two dimensional materials. Adsorption energy is calculated from the output value from GPAW code.

Atomic model can be built by "Atomic Simulation Environment(ASE)", Python module for setting up, manipulating, running, visualizing and analyzing atomistic simulations[84, 85].

4.1.1 Installation

GPAW is available on Linux and MacOS. In this study, GPAW on Linux operating system(Ubuntu 16.04 LTS) is used. Calculation environment was build on the following computation machines:

- 16 GB memory with AMD Opteron(tm) Processor 6164 HE CPU(48 core; 1.7 GHz) $\times 5$
- Project (L) server provided by Hokkaido University academic cloud (10 core; Virtual machine).

To increase the calculation performance, the following packages are preferred to be installed. Those packages are for parallel computation and routine solving system that support DFT calculation, such as matrix calculation.

- **OpenMPI(Message Passing Interface)**
  standard library for parallel computation[86].
- **FFTW(Fastest Fourier Transform in the West)**
  Software library for computing the discrete Fourier transform in one or more dimensions, of arbitrary input size, and of both real and complex data[87].
- **LAPACK(Linear Algebra PACKage)**
  Software library that provides routines for solving systems of simultaneous

linear equations, least-squares solutions of linear systems of equations, eigenvalue problems, and singular value problems[**?**].

- **ScaLAPACK(Scalable Linear Algebra PACKage)**
  Software library of high-performance linear algebra routines for parallel distributed memory machines[88].

Those required packages are easily to install by using APT(Advanced Package Tool), a default software managing tool on Ubuntu. In particular, "apt-get" command is used for the installation by typing the following command in terminal.

- sudo apt-get update (updating software source libraries)
- sudo apt-get install gpaw (installing GPAW code with related programs, such as Python-ASE, libxc, etc)
- sudo apt-get install python-matplotlib (installing visualization libraries for Python, used as a part of Python-ASE)
- sudo apt-get install python-tk (Needed if GUI environment is used, such as Python-ASE, matplotlib, etc)
- sudo apt-get install liblapack-dev (installing LAPACK)
- sudo apt-get install openmpi-bin (installing OpenMPI)
- sudo apt-get install libopenmpi-dev (installing OpenMPI)
- sudo apt-get install libscalapack-mpi-dev (installing ScaLAPACK)

4.1.2 Applying calculation

The calculation with GPAW is applied in the calculation server on remote control. OpenSSH[89] is used to access for the remote control.

- sudo apt-get install openssh-server (installing OpenSSH server)

Calculation parameter is set by writing Python code(See the tutorial page for detailed informations[90]) Parallel calculation can be submitted by the following command:

- nohup mpirun -np *X* gpaw-python run.py & (*X* is replaced by number of CPU cores operated in calculation)

Note that Bader charge analysis is needed for external program from Henkelam Group at Texas University[87].

Chapter 5

Boron based two dimensional materials

Boron based novel two dimensional materials(B-As, B-Sb, and B-Bi) are discovered and investigated by first principle calculation. The idea of this materials comes from the existence of hexagonal boron nitride(h-BN) and prediction of hexagonal boron phosphide(h-BP) by first principle calculation[92]. In these materials, boron(B) is the common element, and nitrogen(N) and phosphorus(P) are the same Group XV elements. Therefore, the existence of B-As, B-Sb, and B-Bi are considered and investigated by first principle calculation, and possible application is suggested.

5.1 Structure and physical property of single layered materials

Computational workflow of optimization of single layered materials is shown in Figure 5.1. Structures are designed by modifying the structure of h-BN, by replacing N for another Group XV elements(B-N, B-P, B-As, B-Sb and B-Bi). Some two dimensional materials are reported to have puckered structure, such as Si, SiGe, GaP, SnSi, AlSb, InAs, InSb honeycomb two dimensional materials[1]. Hence, both plane and puckered structures are relaxed to consider the existence of puckered

structure. Spin polarization is considered in all calculations. Lattice optimization is performed by expanding and shrinking the lattice in 0.5 % and lowest energy structure is taken Structures of designed two dimensional materials are initially optimized within LCAO based calculation[93] in order to reduce computational time. LCAO calculation is performed under the h-grid spacing of 0.18 Å and special K point of $2 \times 2 \times 1$ in Brillouin-zone sampling[94] with exchange correlation of Perdew Burke Ernzerhof (PBE)[69] where periodic boundary conditions are applied in x and y directions.

Final calculation is performed with finite difference wave function with 0.18 Å and special K point of $12 \times 12 \times 1$ in Brillouin-zone sampling[94] with exchange correlation of Perdew Burke Ernzerhof (PBE)[69] where periodic boundary conditions are applied in x and y directions.

Charge transfer on each constituent atom is calculated by Bader charge analysis method[71, 72].

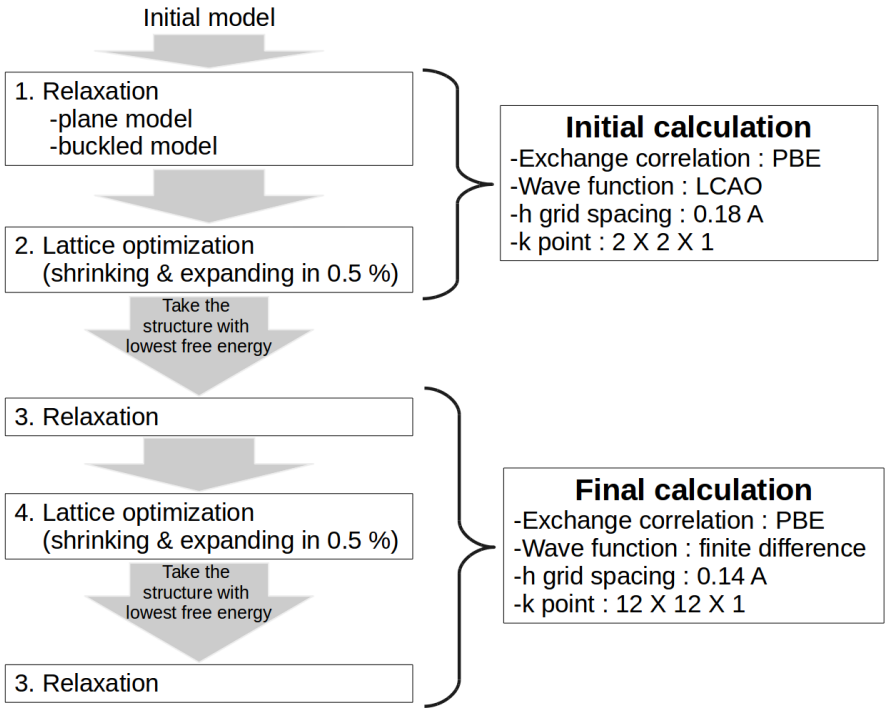

Figure 5.1: Computational workflow for designing novel boron based two dimensional materials.

All optimized single layered structure shows plane honeycomb structure, as

similar as graphene, and h-BN and the unit cell has diagonally 60° and 120° as angles (See Figure 5.2) No puckered structure is found by final relaxation result of all designed materials.

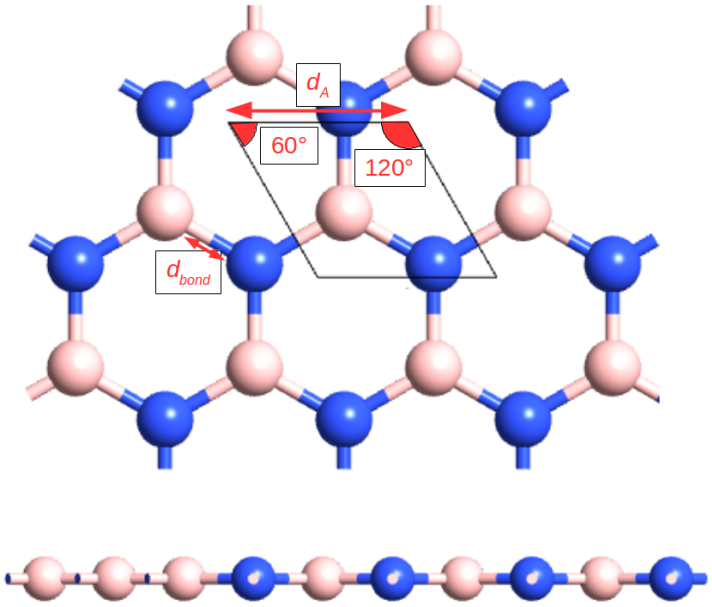

Figure 5.2: Optimized structure of designed two dimensional materials.

The optimized structural details of design materials are collected in Table 5.1. The bond length of two dimensional BN and BP are reported as 1.45 [95] and 1.87 [96]in previous DFT calculation respectively, which are good agreement to optimized bond length shown in Table 5.1. The bond length of h-BN nanosheet was measured experimentally by high-resolution transmission electron microscopy with layer-by-layer peeling of the h-BN nanosheets technique and its value was reported 1.44 Å[97], which is also agreeable to caluclated bond length in Table 5.1. Thus, applied calculation condition is confirmed as sufficient for structural optimization. The lattice constant and distance between B and Group XV element in those materials extend as a period number of Group XV element increases(B-N<B-P<B-As<B-Sb<B-Bi; See Table 5.1). This can be explained as the fact that covalent bond length tends to be longer as the atomic number of elements increases [98, 99, 100].

Table 5.1: Structure property of designed single-layered B-M (M=N,P,As,Sb and Bi) 2D materials.

Materials	B-N	B-P	B-As	B-Sb	B-Bi
d_A[1]	2.51	3.20	3.40	3.74	3.89
d_{bond}[2]	1.45	1.85	1.96	2.16	2.30

Projected density of state(PDOS) is obtained to evaluate the electronic structure of designed two dimensional materials. The overlap of p-orbitals between boron and Group XV elements with each other, as shown in Figure 5.3 This result indicates that bond is formed between B and XV elements in the designed materials, which supports the fact that binding energy between atoms are exothermic reaction(See Table 5.2).

Calculated binding energies of designed materials are all negative(exothermic), indicating that they are energetically stable. Additionally, those binding energies are similar to dissociation energies of existing compounds[101], in other words, these designed materials are possibly synthesized via experiment judging from its energy stability. Charge transfer values calculated by Bader charge analysis are also shown in Table 5.2 in other to evaluate the bond character. Boron in B-N material transfers 2.1 electrons to nitrogen, indicating the strong ironic bonding formation. However the electron transfer boron to XV elements reduces its value as the increase of Group number of the XV elements. This fact is correspondent to binding energy decrease as the increase of Group number of the XV elements. The one notable point on charge transfer from boron is changed minus to plus between "B-N, B-P, B-As" and "B-Sb, B-Bi". This can be explained by the electronegativity difference($\Delta\chi_{B-M}$) of boron and each Group XV element, which is calculated by the following equation:

$$\Delta\chi_{B-M} = \chi_M - \chi_B \tag{5.1}$$

where χ_M and χ_B are Allen electronegativity[102] of each Group XV element and Allen electronegativity[102] of boron respectively. (The values of Allen elec-

[1]Length of optimized lattice constant (Å) (See Figure 5.2)
[2]Distance between boron and Group XV element (Å)

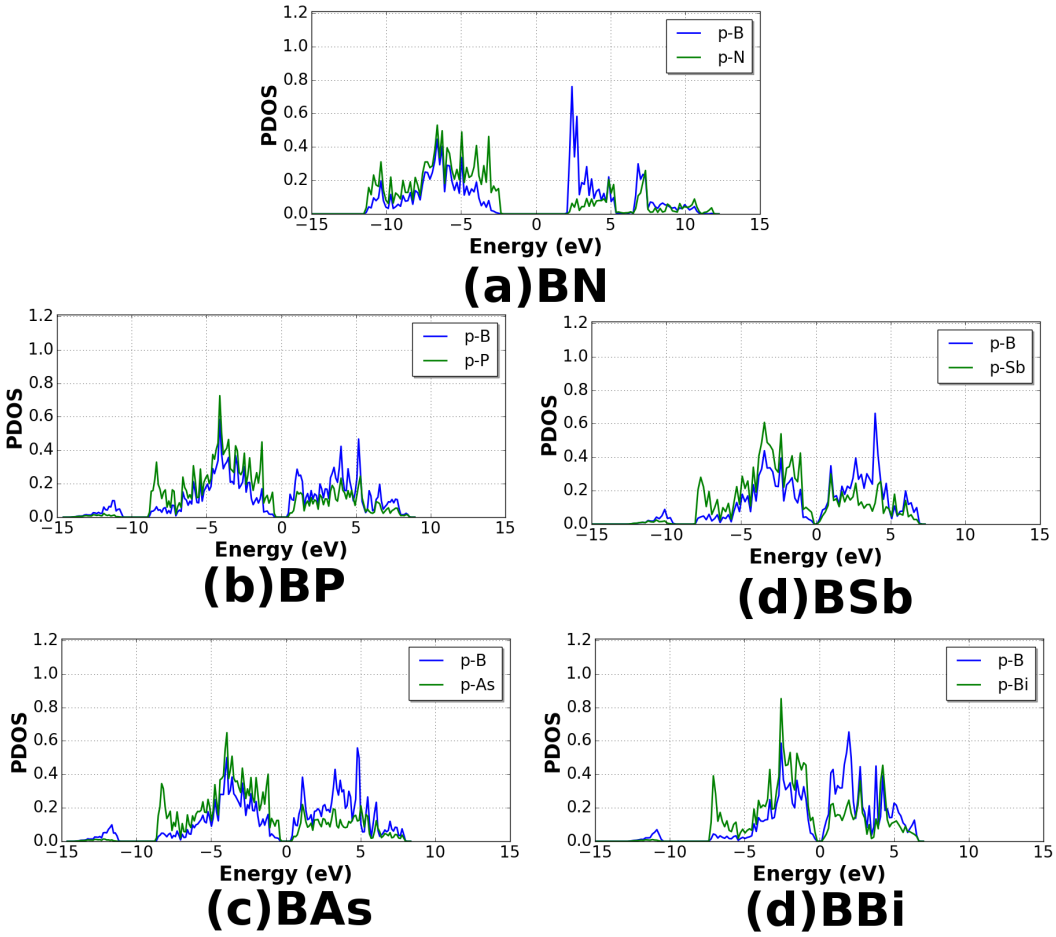

Figure 5.3: Projected density of states(PDOS) of the designed single layer two dimensional materials.

tronegativity of B, N, P, As, Sb, Bi are 2.05, 3.07, 2.25, 2.21, 1.98, and 2.01, respectively.) The electronegativity difference values(See Table 5.2) are also changed minus to plus between "B-N, B-P, B-As" and "B-Sb, B-Bi", as same as the charge transfer from boron(See Table 5.2).

Table 5.2: Physical property of designed single-layered B-M (M=N, P, As, Sb and Bi) two dimensional materials.

Materials	B-N	B-P	B-As	B-Sb	B-Bi
E_{bind}[1]	-3.34	-2.01	-2.76	-1.47	-1.09
E_{gap}[2]	4.00	0.85	0.72	0.23	0.68
V_B[3]	2.13	0.71	0.27	-0.38	-0.26
V_M[4]	-2.11	-0.67	-0.22	0.44	0.334
$\Delta\chi_{B-M}$[5]	1.02	0.202	0.160	-0.067	-0.041

plots the difference in electronegativity ($\Delta\chi_{B-M}$) against the band gaps of the designed materials.

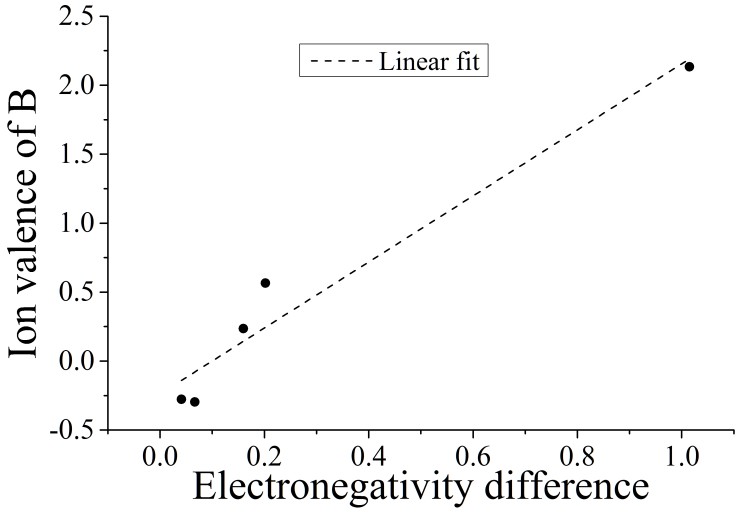

Figure 5.4: The relationship between electronegativity difference ($\Delta\chi_{B-M}$) and band gaps (E_{gap}). Note that the band gap of Group IV - Group Ⅲ-V two dimensional materials are referenced from the paper by Şahin et al.[1].

The electronegativity difference has also strong relationship to band gap of designed two dimensional materials as shown in Figure 5.5. This tendency can be seen another two dimensinal materials, such a as designed Group IV and Group Ⅲ-V two dimensional materials reported in H. Şahin et al.[1], which is Figure 5.6 shows that the electronegativity difference is proportional to the band gap, especially for smaller values. This suggests that the difference in electronegativity can be used as an indicator for predicting the band gap of this type of two dimensional materials.

The reactivity of designed materials are also investigated by the evaluation of absorption energy of hydrogen atom. 2×2 supercell model with hydrogen atom of designed materials are constructed (See Figure 5.7) and first calculation is performed. Two absorption site are considered in the calculation model: on top site of XV elements(See Figure 5.7 (a), and on top site of boron atom (See Figure 5.7

[1]Binding energy between boron and Group XV element (eV/atom))

[2]Band gap (eV) calculated by density of state(DOS) of designed single layer 2D materials(eV)

[3]Charge transferred from boron(B)(eV/Atom)

[4]Charge transferred from M

[5]Electronegativity difference of boron and each Group XV element. The value of electronegativity was applied the one proposed by Leland C. Allen[102]

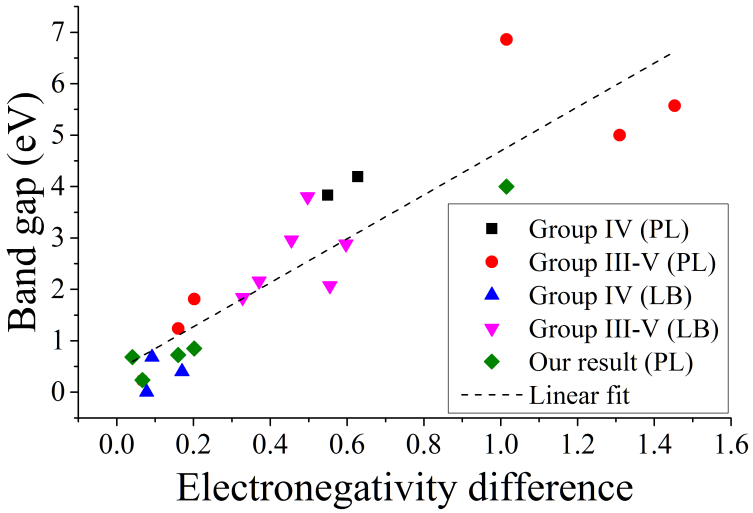

Figure 5.5: The relationship between electronegativity difference ($\Delta\chi_{B-M}$) and band gaps (E_{gap}). Note that the band gap of Group IV - Group Ⅲ-V two dimensional materials are referenced from the paper by Şahin et al.[1].

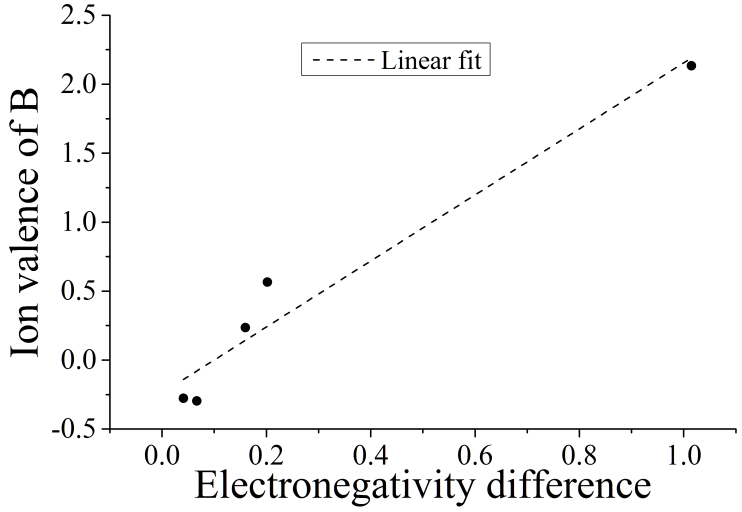

Figure 5.6: The relationship between electronegativity difference ($\Delta\chi_{B-M}$) and charge transfer of boron (δ_{ion})

(b)) To consider the effect of van der Waals force, two different exchange correlation functional is applied for the calculation for comparison: PBE[69] and van der Waals density functional(vdW-DF)[103].

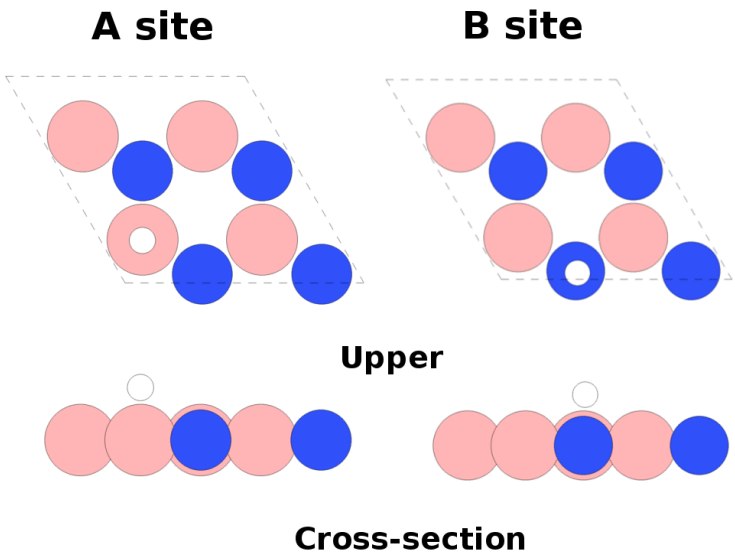

Figure 5.7: The calculation models to evaluate reactivity with H. Pink atoms and blue atoms were indicated as boron and Group XV elements respectively. The one which H locates on top of B named "A site", the other which H locates on top of M (=N, P, As, Sb, Bi) is named "B site".

The calculated adsorption energy with PBE($E_{ads-PBE}$) and vdW-DF($E_{ads-vdW}$), and the distance between hydrogen atom and nearest atom with PBE(d_{PBE}) and vdW-DF(d_{vdW}) are shown in Table 5.3. The adsorption energies in Site 2 show positive value(endothermic) both PBE and vdW-DF functional calculated results, indicating no reactiveness. On the other hand, the adsorption energy in Site 1 shows positive value(exothermic) in B-Sb and B-Bi, with adsorption energies -0.20 eV and -0.55 eV in PBE functional, and -0.25 eV and -0.05 in vdW-DF functional, respectively. The binding energies of B-Sb and B-Bi are weaker than those of B-N, B-P, and B-As, which would allow for greater reactivity in the presence of other atoms. In general, defect free two dimensional materials such as graphene and boron nitride are not reactive against hydrogen, however, defect free two dimensional B-Sb and B-Bi are reactive against hydrogen. Thus, reactive defect free two dimensional materials are discovered. The distances between hydrogen atom and nearest atom are no significant difference in each material calculated result in

both PBE and vdW-DF functional. This fact can be understood that van der Waals force has no effect when calculating the adsorption energy nor the hydrogen-surface distance.

Table 5.3: The calculation result of 2×2 supercell model with hydrogen atom for reactivity evaluation.

Site	Materials	Nearest atom from H	$E_{ads-PBE}$[1] (eV)	$E_{ads-vdW}$[2] (eV)	d_{PBE}[3] (Å)	d_{vdW}[4] (Å)
Site 1	B-N	B	2.30	2.44	1.34	1.32
	B-P		0.61	0.70	1.22	1.21
	B-As		0.23	0.24	1.20	1.20
	B-Sb		-0.20	-0.25	1.20	1.20
	B-Bi		-0.55	-0.05	1.20	1.20
Site 2	B-N	N	3.05	3.20	1.08	1.09
	B-P	P	0.84	0.95	1.43	1.43
	B-As	As	1.05	1.16	1.54	1.54
	B-Sb	Sb	0.87	0.96	1.72	1.73
	B-Bi	Bi	1.19	1.32	1.83	1.86

5.2 Structure and physical property of multi layered materials

Two-layer models are optimized and investigated in order to understand the properties of the B-Group XV element two dimensional materials when layered. Multi-layered are are investigated by as same method as single layer calculation showed in Figure 5.1 with the same cell size structure model. To consider the van der Waals forces within inter-layer sections, both Perdew Burke Ernzerhof (PBE) [69] and van der Waals density functional (vdW-DF)[103] are implemented for the calculation. Two types of layered structure is considered to build starting model for the calculation: “Cross” and “Overlap” structure(See Figure 5.8), which “Cross” structure is general form of hexagonal boron nitride(h-BN).

[1] Adsorption energy calculated with the PBE exchange correlation

[2] Adsorption energy calculated with the vdW-DF exchange correlation

[3] Distance between hydrogen atom and the nearest atom calculated with the PBE exchange correlation

[4] Distance between hydrogen atom and the nearest atom calculated with the vdW-DF exchange

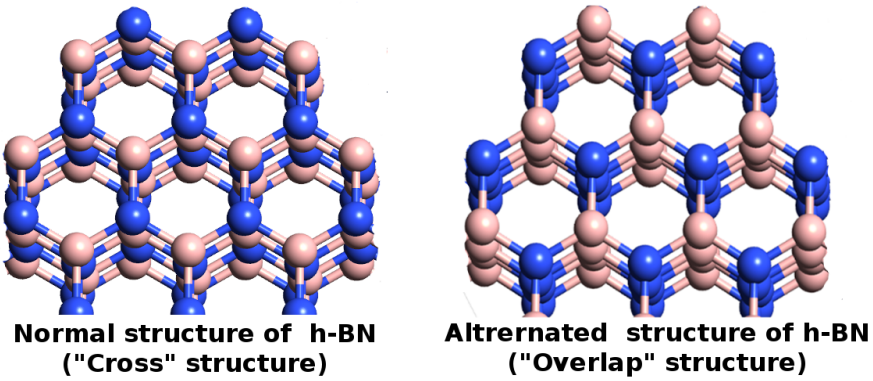

Figure 5.8: An example of normal structure and alternated structure of h-BN, named "Cross" and "Overlap" structure

The optimized structures are showed two types form, "zigzag" and "plane" from the cross section view(See Figure 5.9). The interlayer binding energies are calculated by equation(3.21) and shown in Table 5.4, to evaluate the existence of van der Waals force between layers. The binding energies calculated with PBE functional are positive (endothermic) in B-N, B-P, and B-As, while those of calculated with vdW-DF are negative (exothermic). This fact shows that B-N, B-P, and B-As are formed layered structure by weak physisorption between interlayer, which is originated from van der Waals force. The binding energy of B-As in "zigzag" structure"(See Figure 5.9) is positive (endothermic) for both PBE and vdW-DF functional calculations, indicating that "zigzag" structure of B-As is energetically unstable.

Distance between the closest atoms between facing layers (d_{near}) is calculated structure by vdW-DF functional, as shown in Table 5.5. d_{near} of B-N in "Cross" structure is calculated as 3.559 Å(See Table 5.5), which is in good agreement with the previously-reported experimentally measured distance of 3.33 Å[104]. d_{near} of B-N and B-P in "Overlap" structure are slightly longer than those of "Cross" structure. B-As, meanwhile, is seen to have an interlayer distance of 1.814 Åwhen in a "Cross" structure, the shortest of the reported "Cross" structures, and an interlayer distance of 4.379 Å, when in an "Overlap" structure, the longest of the reported

correlation

[1]Interlayer binding energy between layers calculated with the PBE exchange correlation

[2]Interlayer binding energy between layers calculated with the vdW-DF exchange correlation

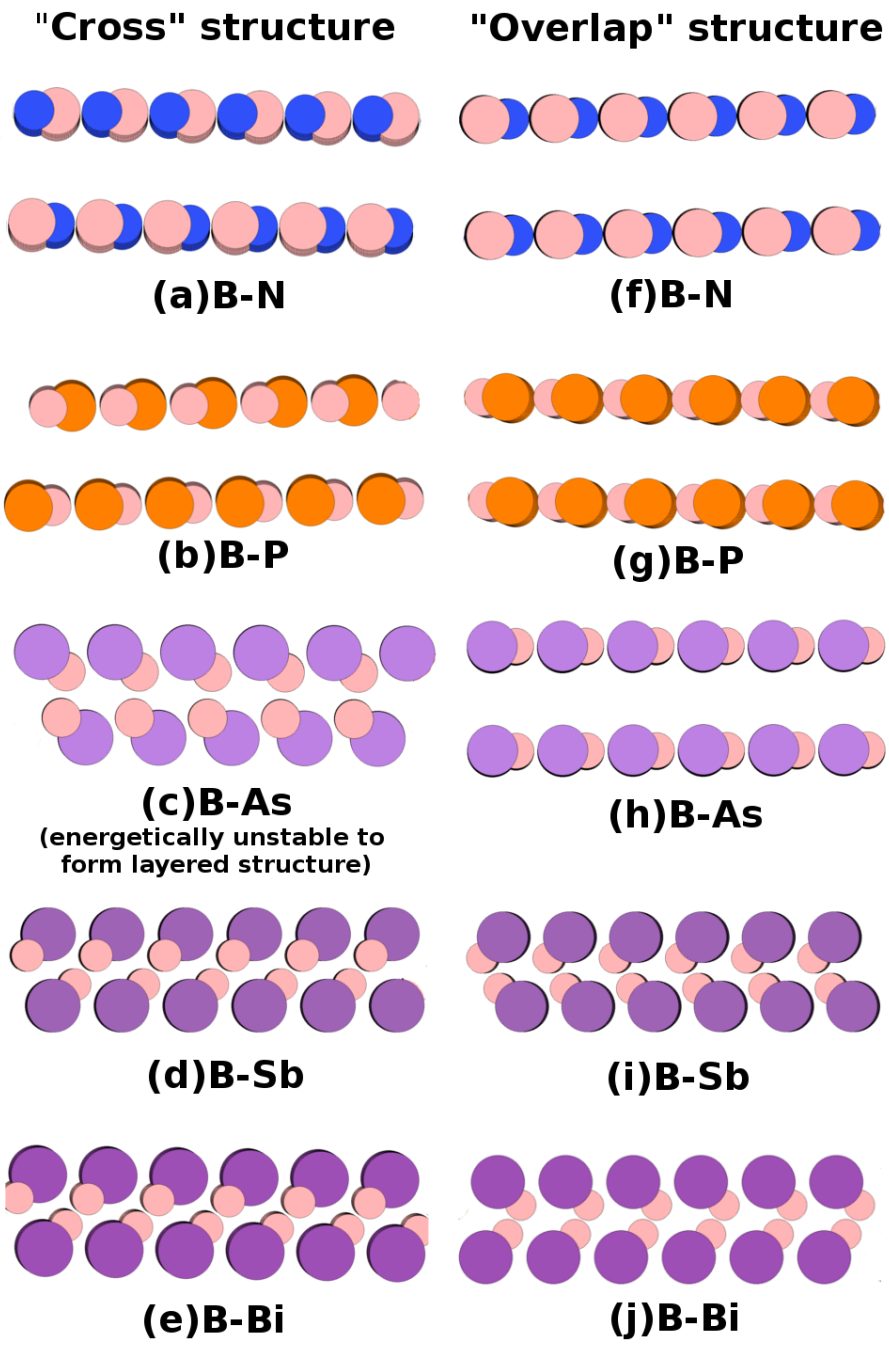

Figure 5.9: Cross section view of designed 2 layered designed two dimensional materials.

Table 5.4: Interlayer binding energy between layers(E_{layer}).

Materials	Structure	$E_{layer-PBE}$[1](eV)	$E_{layer-vdW}$[2](eV)
B-N	Cross	5.59×10^{-2}	-4.51×10^{-2}
	Overlap	4.50×10^{-2}	-9.13×10^{-2}
B-P	Cross	1.55×10^{-2}	-12.0×10^{-2}
	Overlap	2.49×10^{-2}	-9.66×10^{-2}
B-As	Cross	0.80×10^{-2}	17.5×10^{-2}
	Overlap	1.73×10^{-2}	-9.84×10^{-2}
B-Sb	Cross	-1.22	-1.17
	Overlap	-0.85	-0.83
B-Bi	Cross	-2.04	-2.07
	Overlap	-1.67	-1.68

"Overlap" structures.

Table 5.5: Distance between the closest atoms between facing layers (d_{near}) from the results using exchange correlation function of vdW-DF

Materials	Structure	d_{near} (Å)
B-N	Cross	3.559
	Overlap	3.649
B-P	Cross	3.918
	Overlap	4.176
B-As	Cross	1.814
	Overlap	4.379
B-Sb	Cross	2.636
	Overlap	1.710
B-Bi	Cross	2.737
	Overlap	1.619

Charge transfer on the upper layer is evaluated by Bader charge analysis(See Table 5.6). The charge transfer values on boron(B) are positive in B-N, B-P, and B-As, while the those of B-Sb and B-Bi are negative, as similar as the single layer case as shown in Table 5.2

The reason of structure formation of "plane" and "zigzag" (See 5.9) can be provided to consider the interlayer bonding states in each materials. The interlayer binding energy between layers of "overlap" B-N, B-P, and B-As is considered as

[1]Charge transferred from boron(B)(eV/Atom)
[2]Charge transferred from Group XV elements(*M*)(eV/Atom)

Table 5.6: Charge transferred from B to *M* in the designed 2 layered B-*M* (M=N,P,As,Sb and Bi) materials calculated with exchange correlation function of vdW-DF. Note that the charge transfer is calculated by Bader analysis on the upper layer.

Materials	Structure	V_{layerB}[1]	V_{layerM}[2]
B-N	Cross	2.12	2.11
	Overlap	2.09	2.08
B-P	Cross	0.86	0.84
	Overlap	0.74	0.72
B-As	Cross	0.005	0.033
	Overlap	0.29	0.27
B-Sb	Cross	0.49	0.50
	Overlap	0.40	0.46
B-Bi	Cross	0.47	0.47
	Overlap	0.34	0.44

physisorption originated from van der Waals force due to their values calculated from vdW-DF functional result(See 5.4). On the other hand, the interlayer binding energy between layers of "zig-zag" B-Sb and B-Bi is the reflection of chemisorption due to their their values calculated from PBE functional result(See Table 5.4). The charge densities of two layered B-Sb and B-Bi (See 5.6) is smaller than those of single layer B-Sb and B-Bi(See Table 5.2). This fact show that those two layered materials form an ionic bond in the interlayer between B and Sb or Bi each other(See Figure 5.10), which supports the forming of chemisorption occurrence in "zig-zag" B-Sb and B-Bi. Additionally, the Sb and Bi atoms have a larger atom radius than B, which could be changing the "plane" form to the "zigzag" form in each layer.

Density of states(DOS) are investigated to investigate electronic structure of two dimensional materials, as shown in Figure 5.11 and Figure 5.12 DOS of bilayer two dimensional material is strongly coupled when layers are constructed as shown in Figure 5.11 ("Cross" case) and Figure 5.12("Overlap" case). The peaks as well as band gaps are significantly different. This suggests that electronic structure of bilayer two dimensional material can be essentially tuned and design by controlling the structures of layers.

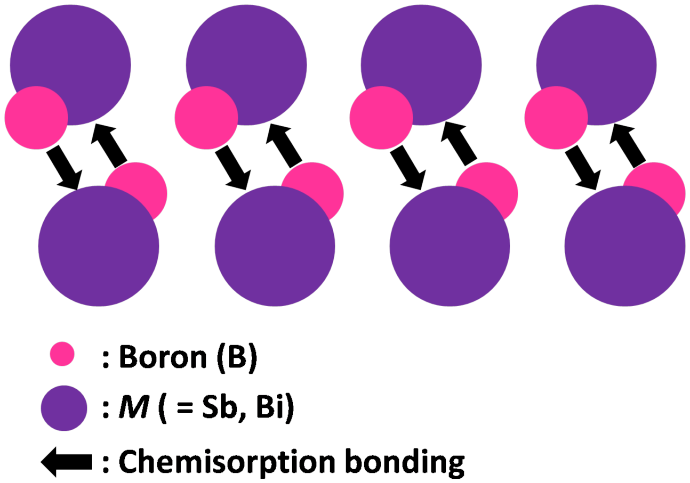

Figure 5.10: The scheme of the reason of “Zigzag” formation.

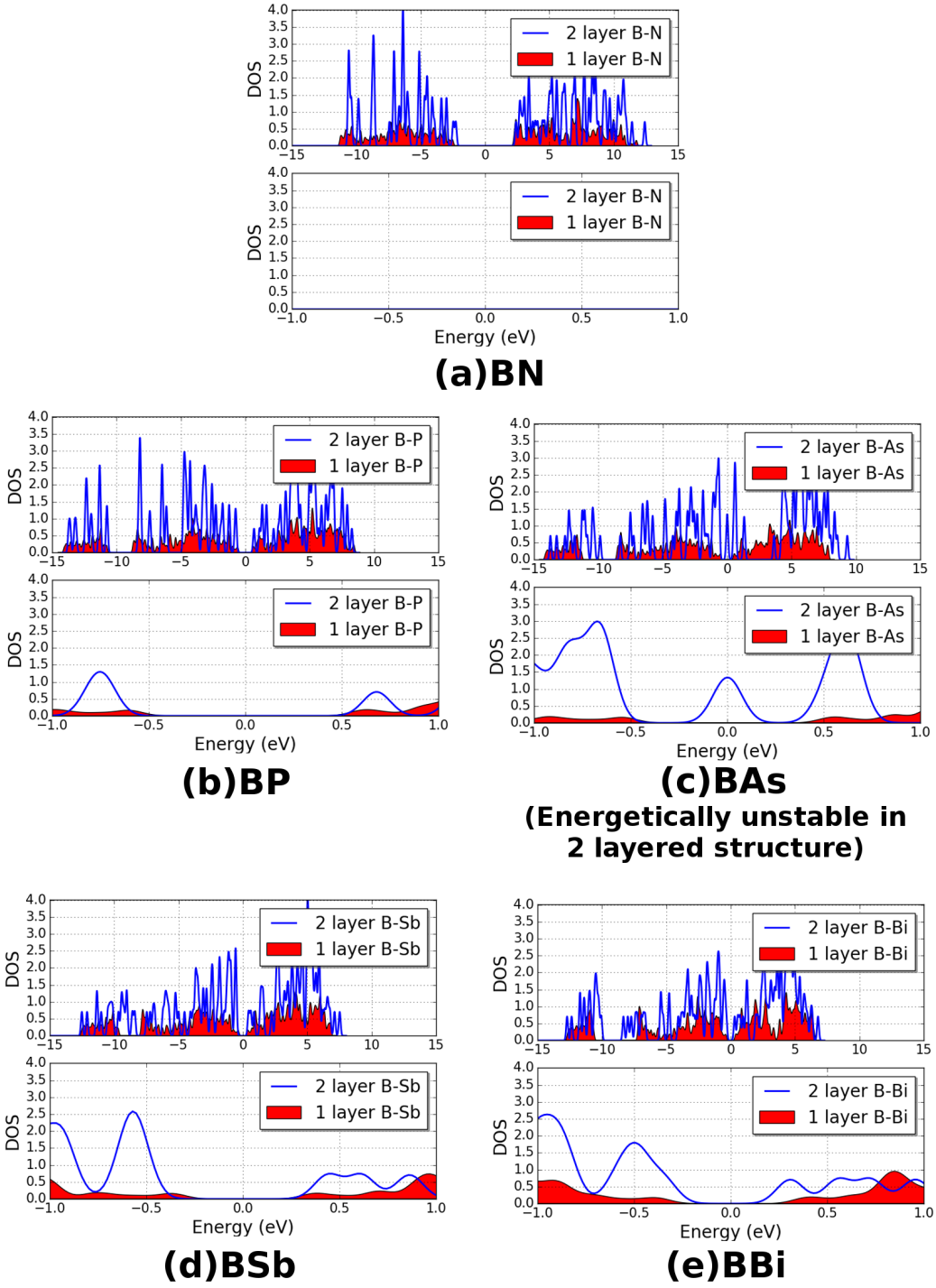

Figure 5.11: Density of states (DOS) of designed 1 layered (filled) and 2 layered (plotted) materials relaxed from "Cross" structure

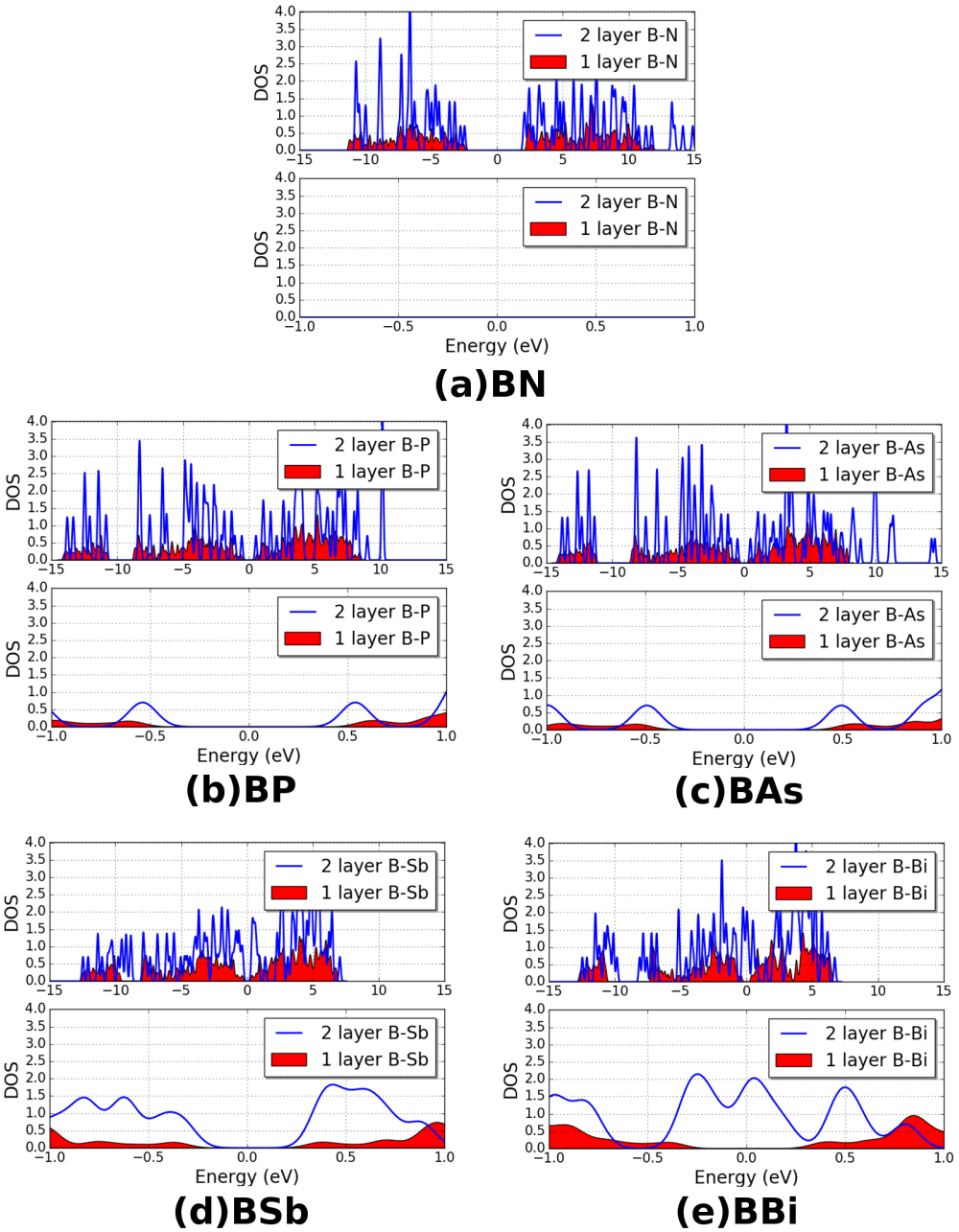

Figure 5.12: Density of states (DOS) of designed 1 layered (filled) and 2 layered (plotted) materials relaxed from "Overlap" structure

Chapter 6

A method of Materials Informatics

Materials Informatics approach in this study is Machine leaning is subset field in the academic field of Computer Science, aimed to “Artificial Intelligence(AI)” , which aims to realize “learning” process by computer. The word “machine learning” was made by Arthur Samuel[105]. The machine learning consists of various algorithm to recognize statistic “rule” in given dataset. Those machine learning algorithms are roughly classified by 2 categories as the follow.

Supervised learning Supervised learning is the learning method with given two (“input(target)” and “output(descriptor)”) pair of datasets. Computer is “supervised” between these two “input(target)” and “output(descriptor)” by applied learning algorithm. This learning method can be divided in two groups due to the types of applying data: “Regression” and “Classification”. “Regression” is applied to the data that express “value”, in other words, “Regression” can handle all physical parameter without data preprocessing. On the other hand, “Classification” can only be applied to the data that express “class”, for example, “oxide” or “non-oxide”, “conductive” or “insulative” etc. However, “Classification” can be applied to “value” by classifying.

Unsupervised learning Unsupervised learning is the learning algorithms to discover structures and properties from the learning data without answers. This learning method can be grouped in two main algorithms: “Clustering” and “Association”. Clustering aims to finds the inherent groups in the data, for

example, finding which elements in the substance is effective to the stability. Association aims to discover the rule that explains large portions.

In this study, supervised learning is generally used to predict novel materials since the property target is specified when the materials designed by Materials Informatics approach. Thus, this section deals with how supervised learning is applied for actual material exploration.

- **Data preparation** Data must be prepared in advance for machine learning, by obtained from both theoretical and experimental result. Those data is normally collected in database for unitary management, under standard format and unified unit system. In Informatics approach, negative data has great meaning because data is processed statistically. Therefore, failure data should be dealt with as same as successful data that tend to be reported in publication.

- **”data preprocessing“, ”descriptor hunting“, and ”algorithm choice“** Before machine learning is applied, 3 important process is applied to obtain accurate material prediction: ”data preprocessing“, ”descriptor hunting“, and ”algorithm choice“(Figure 6.1) **Data preprocessing**
  In a many cases, collected data is not directly applicable for machine learning. For example, unit standardization in data, classifying target data in order to implement classifier algorithm in machine learning, etc. On another sight, data addition is performed to help search descriptors, such as adding electronegativity of constituent atom. Data preprocessing is applied appropriately for successful machine learning **Descriptor hunting**
  Descriptor is a dataset(physical value etc.) which describes ”target“ in statistic aspect. The important rule of choosing descriptors is ”reversible“. In solving ”inverse problem“ process, the descriptor would be the input to predict ”target“ value, hence, it is important to choose descriptors that can estimate what material is. For example, atomic number of constituent atom can be good descriptor because it can provide the composition of materials from itself. On the other hand, for example, it is difficult to estimate material from

thermal conductivity, therefore, it is unsuitable physical property for descriptor **Algorithm choosing**

Machine learning algorithm needs carefully to be chosen by judging from train data structure. For example, deep learning algorithm generally requires mass of data(>100,000) so enough quantity of data is needed. The important thing is to choose appropriate algorithm which can capture data structure correctly, not increasing fitting degree.

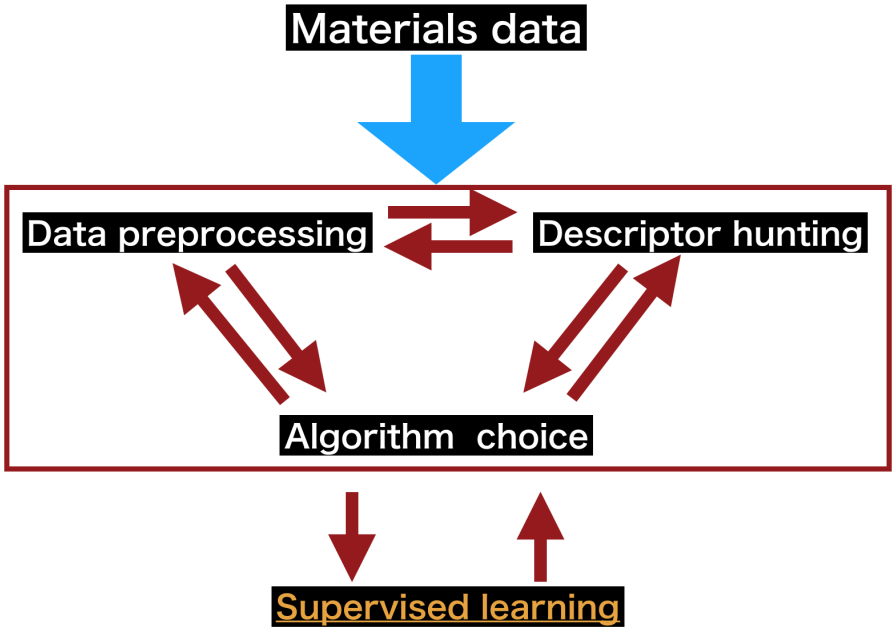

Figure 6.1: The scheme of supervised machine learning.

- **Machine learning**

  For the novel materials prediction, supervised machine learning is the most common way(Figure 6.2). Firstly, target material property for designing materials(e.g. higher magnetic moment, particular band gap etc.) must be set for ”answer“ for machine learning. Then, descriptors are chosen by the reference from training material data. For example, the structure of materials, atomic number of constituent atoms, etc. In the machine learning process, computer learns those paired ”descriptors(input)“ and ”target property(answer)“. By this learning process, computer read the data structure and get ”smart“. Trained data can be evaluated by “cross validation”(See next section “Cross

validation"). By judging from cross validation score, evaluation of descriptors, etc., machine learning model is set for next step: solving "Inverse problem".

- **Inverse problem**

  The extracted descriptors are input into the "smart" computer and inverse problem is solved. Consequently, the "smart" computer predicts target materials property from each extracted descriptors. Then, preferable material property(e.g. higher magnetic moment, particular band gap etc.) is extracted from whole predicted list, then, the extracted list would be the potential material candidates(Figure 6.2).

- **Evaluation**

  The candidates are statistically predicted in principle, which has no relationship with actual material property in themselves. Thus, evaluation of candidates is needed ether theoretical or experimental way. As seen in the result of evaluation, actual material candidates are revealed, and it can be feedback to add lacking data in material database in order to increase accuracy of prediction as necessary.

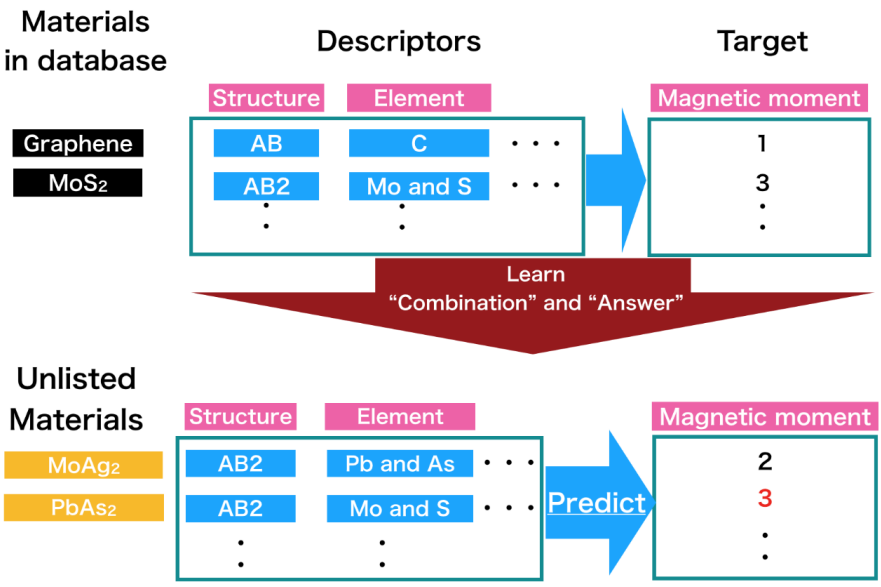

Figure 6.2: The scheme of supervised machine learning.

On the other hand, however, unsupervised learning can be useful to capture the data decline within materials database for descriptor hunting.

6.1 Cross validation

Cross validation is a method to evaluate the learning algorithm and descriptor choice in supervised learning. k-fold cross-validation is popular and it performed by splitting two data blocks: "test" data(K) and "train" data($K-1$)(Figure 6.3) in k-order.

Each k(=1,2,..,K) fits the parameter λ(subset size, etc) to the other $K-1$ parts. Then, the error in predicting kth part can be calculated by giving $\hat{\beta}^{-k}(\lambda)$ fitted function:

$$E_k(\lambda) = \sum_{i \in kth\ part} (y_i - xi\hat{\beta}^{-k}(\lambda))^2 \tag{6.1}$$

Therefore, the cross validation error can be expressed as the following equation:

$$CV(\lambda) = \frac{1}{K}\sum_{k=1}^{K} E_k(\lambda) \tag{6.2}$$

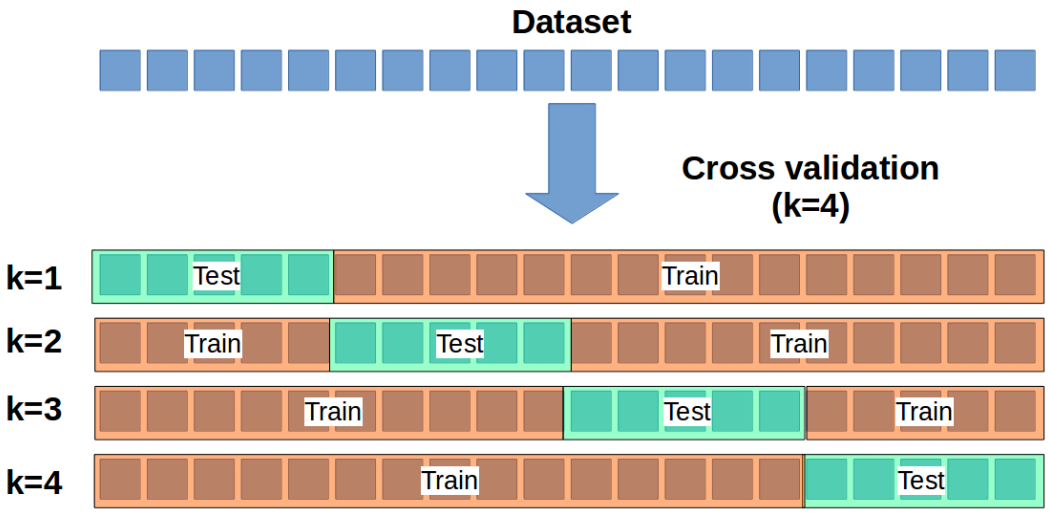

Figure 6.3: The scheme of k-fold cross-validation(k=4).

Note that cross validation is performed within certain given data for machine learning. Hence, it can evaluate the degree of fitting in given data dimension, however, the evaluated score has nothing to do with external dimension of data.

6.2 Naive bayes

Naive Bayes is supervised learning algorithm based on Bayes' theorem. This algorithm is well known for the application to text categorization, such as spam filter in e–mail. In this study Gaussian Naive bays classifier, he derived model of Naive Bayes classifier that conditional probability($P(A|B)$) is assumed as Gaussian distribution, is implemented for machine learning in the consideration of data structure of trained data. Bayes' theorem is defined as the following equation (6.3)[106].

$$\frac{P(A|B) = P(B|A)P(A)}{P(B)} \tag{6.3}$$

- $P(A|B)$: Conditional probability

- $P(A)$: Probability of observing A

- $P(B)$: Probability of observing B

where A and B are event and $P(A) > 0$.

As the classifier model in conditional probability, the model is defined as equation (6.4)[107]:

$$p(C_k|x_1,...,x_n) \tag{6.4}$$

for each K possible results or class C_k where a vector x=$x_1,...,x_n$ with n independent features.

$$p(C_k,x_1,...,x_n) \tag{6.5}$$

As considering equation (6.4) and (6.5) the equation is rewritten as the following equation (6.9) by using chain rule for for repeated application.

$$p(C_k, x_1, ..., x_n) = p(x_1, ..., x_n, C_k) \tag{6.6}$$

$$= p(x_1|x_2, ..., x_n, C_k)p(x_2|x_3, ..., x_n, C_k) \tag{6.7}$$

$$= ... \tag{6.8}$$

$$= p(x_1|x_2, ..., x_n, C_k)p(x_2|x_3, ..., x_n, C_k)...p(x_{n-1}|x_n, C_k)p(x_n|C_k)p(C_k) \tag{6.9}$$

Assuming that each feature x_i is independent other future $x_j (i \neq j)$ in given the category C_k, the following equation (6.10) is introduced:

$$p(x_i|x_{i+1}, ..., x_n, C_k) = p(x_i|C_k) \tag{6.10}$$

Thus, joint model is introduced as equation (6.11)

$$p(C_k|x_1, ..., x_n) \propto p(C_k, x_1, ..., x_n) \tag{6.11}$$

Therefore, equation can be solved as follows:

$$= p(C_k)(x_1|C_k)p(x_2|C_k)p(x_3|C_k)... \tag{6.12}$$

$$= p(C_k)\prod_{i=1}^{n} p(x_i|C_k) \tag{6.13}$$

To construct classifier from the probability model, a class label $\hat{y} = C_k$ is introduced. Thus, the Naive Bayes classifier is defied as follows:

$$\hat{y} = \underset{K \in \{1,...,K\}}{\operatorname{argmax}} \; p(C_k)\prod_{i=1}^{n} p(x_i|C_k) \tag{6.14}$$

Chapter 7

Implementation of machine learning environment

Machine learning(ML) environment is now provided by various form. In the commercial software, MATLAB supports machine learning environment by applying the optional module " Statistics and Machine Learning Toolbox". It also provides IoT solutions, such as integration to Raspberry Pi, and used by various companies. Therefore, it provides high efficiency to develop integrated system for data-driven solutions. R language is the most popular programming language among researchers for statistical computing, including modules for machine learning. It is also provided under GNU General Public License so the license. However, Python language is getting popular for data scientific solutions recently. Python itself is a high-level programming language for general-purpose programming; however, the various modules are implemented, which includes data processing, machine learning and data handling. The representative libraries for machine learning are below:

- **Theano**
  Library for deep learning primarily developed by a Montreal Institute for Learning Algorithms at the Université de Montréal[108]

- **TensorFlow**
  Machine learning library for deep learning developed by Google under the "Google Brain" project[109].

- **Caffe**

  Framework for image processing based on constitutional neural network(CNN) implemented in MATLAB[110]

- **Chainer**

  Framework for neural network processing developed by Japanese venture company Preferred Networks[111] in partnership with IBM, Intel, Microsoft, and Nvidia[112].

- **Scikit-learn**

  Machine learning library for Python language[113].

Each library has own characteristic aspects in its functions(GPU support, algorithms, code readability, etc), thus, developers can choose the appropriate library for their own demands for implementation.

Python also has various libraries for scientific analysis. The representative (and used in this study) modules are introduced below:

- **Numpy**

  Python modules for advanced calculation, such as many dimensional array(vectors, matrix etc) calculation. The core program is written with C or FORTRAN language in order to accelerate computation performance[114].

- **Scipy**

  Python library for numerical analysis, such as image processing, signal processing etc[115].

- **Pandas**

  Python library for data manipulation and analysis[116].

- **Matplotlib**

  The visualization modules for Python[117].

Those libraries can be integrated in one code by importing each modules. For example, data preprocessing for Pandas, machine learning for Sickit-learn, and vi-

sualization for Matplotilib. This flexibility of integration makes Python popular in data scientist.

7.1 Scikit–learn

Scikit–learn is a python library for data mining and data analysis built under BSD license[113]Scikit–learn supports various kind of machine learning algorithms: regression, classification and clustering. Scikit–learn is mainly written in Python language so it is easy to integrate with another Python libraries, such as Scipy (data processing), Seaborn(data visualization), etc. Some core algorithms are build with Cython to improve computational performance since Python itself performs comparably slow because it is script language. Due to its abundant option of data analysis tools and multiplicity of use, a lot of commercial companies are also used to build their own platforms, such as “Sportify”, “Evernote”, etc[118].

7.1.1 Functions

Scikit-learn supports various kind of machine learning algorithms. The version of 0.20.2 supports the following functions below[119]:

- **Classification**

  Algorithms to identify to which category an object belongs to, such as linear models, kernel ridge regressions, support vector machines(SVM), nearest neighbors, Naive Bayes, ensemble methods, etc.

- **Regression**

  Algorithms to predict a continuous-valued attribute associated with an object, such as linear models, kernel ridge regressions, support vector machines(SVM), nearest neighbors, Naive Bayes, ensemble methods, etc.

- **Clustering**

  Algorithms to group of similar objects into sets, such as K-means, affinity propagation, mean-shift, spectral clustering, Gaussian mixtures, etc.

- **Dimensionality reduction**

Algorithms to reduce the number of random variables to consider, such as principal component analysis(PCA), future selections, etc.

- **Model search**

  Algorithms to compare, validate and choose parameter and models, such as grid search, cross validation etc.

- **Preprocessing**

  Algorithms for data preprocessing, such as feature extraction and normalization.

In the tutorial page of Scikit-learn provides the algorithm cheat-sheet to help choosing algorithm[120] within the implemented various types of algorithm.

7.1.2 Installation

Scikit-learn can be installed with various way. In this study, programming environment is installed on operating system(Ubuntu 16.04) by using APT(Advanced Package Tool), a default software managing tool on Ubuntu.

To install Scikit-kearn by APT, the following command is typed and executed on terminal:

- sudo apt-get install python-sklearn
- sudo apt-get install python-sklearn-pandas

Note that version of Scikit-learn is older(0.17.0-4) when you installed with APT on Ubuntu 16.04(The newest version is 0.20.2 on December 2018).

There are two another installation ways for Scikit-learn: by using “PIP” and “Anaconda”, which way is available for new version of Scikit-learn. PIP is package management tool for Python which is available under Python development environment. Scikit-learn is installed by using command “pip install -U scikit-learn”.

Anaconda is a data science platform distributed by Anaconda, Inc[121]. Both PIP and Anaconda are free and open-source, cross-platform suitable for Windows, Linux and MacOS. For the installation, binary file are available on official website[121]. Anaconda initially contains Scikit-learn.

Note that the explained 3 installation ways are not mixed in one computer unless the package dependencies of Python would be corrupted and result to program error.

Chapter 8

Two dimensional magnets

Novel two dimensional magnet materials are discovered by Materials informatics approach. The idea of this study comes from the recent achievement of material discovery by data science with first principle calculation datasets[123, 124, 125, 126]. First principle calculation is commonly used method on computational physics, and several new materials are designed, such as boron nitride nanotubes[127], post-perovskite phase of $MgSiO_3$[128] and so on. High-throughput calculation is one of the efficient ways to obtain large amount of homogeneous data[129] to apply data science method and some computational database are available on the Internet, such as AFLOWLIB.ORG[130], COMPUTATIONAL MATERIALS REPOSITORY[131] and Materials project[132] In this study, two dimensional magnet materials are designed by using materials informatics approach since it is possible to synthesize two magnet materials experimentally as the one example Co and chromium triiodide (CrI_3) films[133, 134] for application, such as electrical control for realizing magnetoelectronics[135].

8.1 Method

The procedure of Materials informatics method of this study is mainly classified in 3 steps: (1)Machine learning, (2)Reversed problem, and (3)Evaluation. However, the most of cases, it is rare that prepared data is simply applicable to the machine learning. Hence, data preprocessing is applied, for example, adding extra infor-

mation(atomic number, electronegativity of constituent atoms, etc) for searching appropriate descriptors. The preprocessing data itself is performed as necessary for machine learning, such as normalization, classification of data, etc.

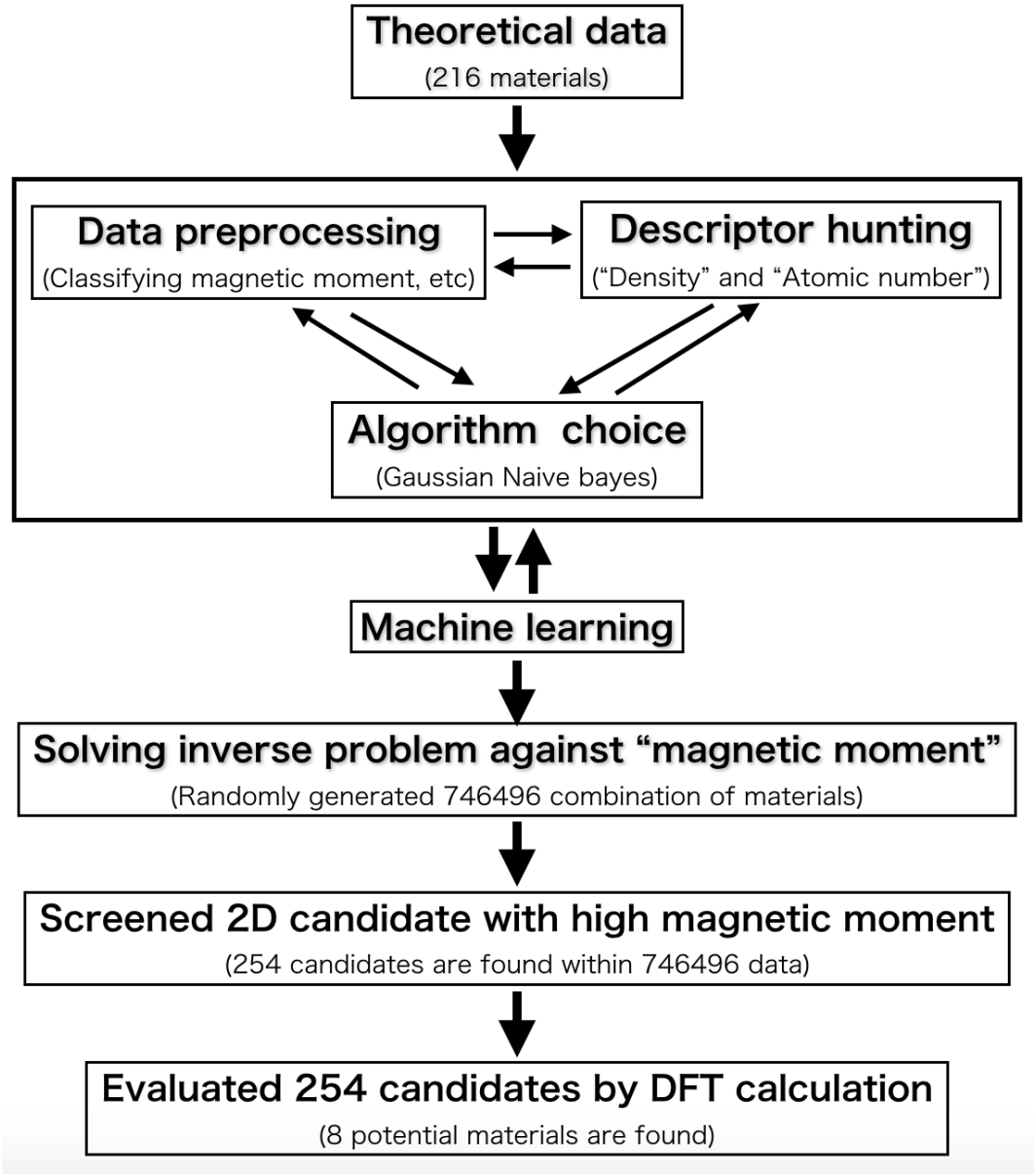

Figure 8.1: The overview of exploring two dimensional magnets by Materials Informatics approach

Applied data is obtained from the computational material database published on the Internet “Computational 2D Materials Database”[2] as a part of “COMPUTATIONAL MATERIALS REPOSITORY”[131] 216 two dimensional materials data(Note that this number is varied in recent since the data has been uploaded by

the owner) is explored to determine descriptors for the magnetic moments. Magnetic moment values obtained from the database are classified in 4 groups for data processing: 0 μ_B-0.4μ_B as 0 (157 data), 0.5μ_B-1.4$\mu_B$as 1 (23 data), 1.5μ_B-2.4μ_B as 2 (22 data), and 2.5μ_B-3.4μ_B as 3(14 data). This classification is aimed in two informatics aspects: (1) making structural data easier by classifying 4 groups, which enables to apply classifier algorithm in machine learning. (2)Reducing data bias by intentional classification rule(trained data contains relatively lager material data with low magnetic moment. Thus, classification rule is set in narrow range in lower magnetic moment). The trained machine is evaluated using cross validation where the data is randomly organized into 10 % test data and 90 % trained data and average score of 10 random test and trained data set is taken.

In the trained data, there are two structure types of two dimensional materials (MoS_2 based AB_2, and graphene based AB structures), as shown in Figure 8.2. Therefore, the reversed problem is performed to predict novel AB and AB_2 type two dimensional materials with high magnetic moment by using learning algorithm with high cross validation score against trained data(in this case, Gaussian Naive Bayes classification algorithm. In details, see "Result and discussion" section)

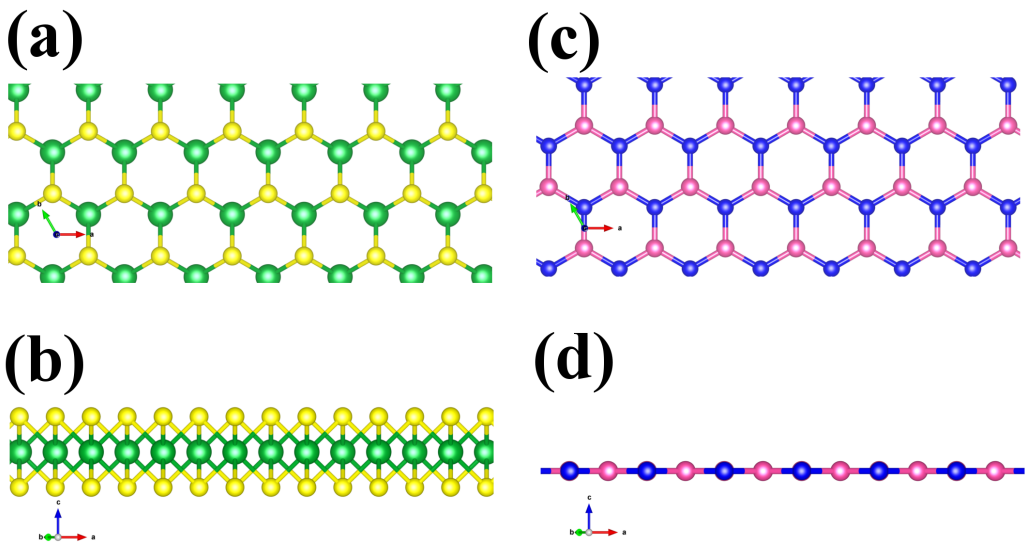

Figure 8.2: The structural models of MoS_2 based AB_2 in top (a) and side (b) view and graphene based AB in top (c) and side (d) view .

Evaluation

Predicted two dimensional materials are optimized in structure and evaluated physical properties by first principle calculation. vdW-DF functional[136] is applied for considering the van der Waals force. Spin polarization is considered in all calculations. Structures of designed two dimensional materials are initially optimized within LCAO based calculation[93] in order to reduce computational time. $4 \times 4 \times 1$ of special k-spacial point Brillouin-zone sampling[94] is set for all calculations where periodic boundary conditions are applied in x and y directions. Lattice optimization is performed by expanding and shrinking the lattice in 0.5 % and lowest energy structure is taken. Magnetic moment (μ_B) per atom and heat of formation (eV/per an atom) from calculated result is plotted in Figure 8.4.

The formation energy per atom of predicted two dimensional materials AB and AB_2 are calculated by Equations 8.1 and 8.2, respectively:

$$E_{AB} = \frac{E_{2DAB} - E_A - E_B}{2} \tag{8.1}$$

$$E_{AB_2} = \frac{E_{2DAB_2} - E_A - (E_B \times 2)}{3} \tag{8.2}$$

where A and B represents the bulk A and B per atom.

8.2 Result and discussion

The data mining is performed to the 216 computational two dimensional materials data[2] and 4 descriptors for predicting magnetic moment is explored: atom number of A element, atom number of B element$\times$X(X is 1 if the structure is AB and 2 if the structure is AB_2), density of A element, and density of B element$\times$X(X is 1 if the structure is AB and 2 if the structure is AB_2). “Gaussian naive bayes classification” is chosen as the appropriate algorithm to predict magnetic moment by the result of the mean score of 73 % in average of 10 random states, the median score of 72%, the standard deviation of 5%, and the highest score of 81 % in cross validation against the provided 4 descriptors. Note that magnetic moment values obtained from the database are classified in 4 groups, explained in “Method” section(0 μ_B-0.4μ_B as

0, 0.5μ_B-1.4μ_B as 1, 1.5μ_B-2.4μ_B as 2, and 2.5μ_B-3.4μ_B as 3). The meaning of these 4 descriptors can be seen by the visualizing the trained 216 data as shown in Figure 8.3 to find relationship descriptors and magnetic moment in order to find tendency within data. As seen in Figure 8.3 (a), high magnetic moments concentrate in the range of 21-25 for atomic number of A(See yellow filled area). Consequently, Figure 8.3 (b) indicates the concentration of high magnetic moments in the range of 6-9 g/cm^3 for the density of atomic number A(6-9 g/cm^3)(See yellow filled area). Those tendency is considered to result to be these 4 descriptors effective for high cross validation score within "Gaussian naive bayes classification" algorithm.

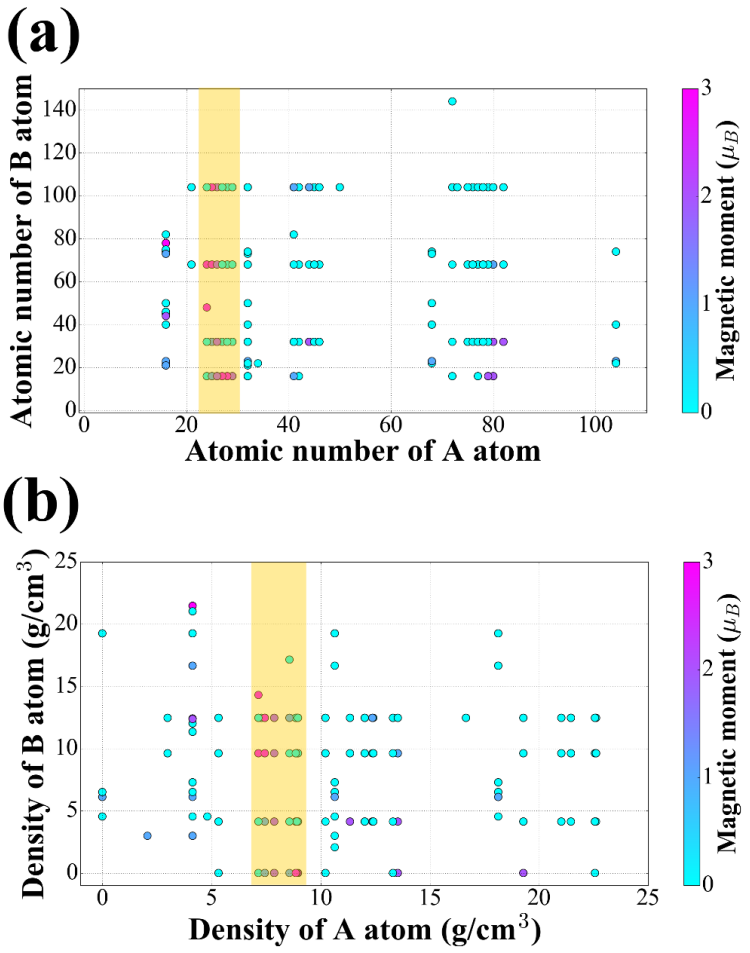

Figure 8.3: (a) Atomic number of A and B with corresponding magnetic moments and (b) Atomic density of A and B with corresponding magnetic moments in reported 216 two dimensional materials data[2]. Note the magnetic moments in data is classified into 4 groups. The area with high magnetic moments are also filled in yellow.

To obtain the combination of composition of novel two dimensional magnets, revered problem is solved with "Gaussian naive bayes classification" algorithm against trained 216 data. The 4 descriptor variables, combination of atomic numbers of 8-82 as well as the corresponding densities are generated to apply reversed problem. For the cases of B_2, the atom numbers and corresponding densities are multiplied by 2. As a result, total of 746,496 predicted material candidates is obtained. Those 746,496 of candidates are screened into the magnetic moment of Group 3(254 candidates), which have a magnetic moment of $2.5 \sim 3.4\dot{9}$. Note that those candidates have no duplication to original 216 two dimensional materials data. Those candidates that might have high magnetic moment are evaluated by DFT calculations for structural optimization, formation energy, and obtaining magnetic moments in each materials, and results are shown in the Table 8.1(AB type two dimensional materials) and Table 8.2(AB_2 type two dimensional materials)

Magnetic moment (μ_B) per atom and heat of formation (eV/per an atom) from calculated result are plotted in the Figure 8.4 in order to discover the potential two dimensional magnets.

The ideal materials need to have stable formation energy and high magnetic moment, therefore, the area of high magnetic moments ($>1.5\mu_B$) and low formation energy(<1 eV) is colored in Figure 8.4. Within the colored area, stable 7 two dimensional materials with high magnetic moment are discovered: $MnPd_2$, FeS, CrSe, CrS, MnTe, MnSe, and MnS. Those materials are not listed original 216 two dimensional material data so it is proven that novel two dimensional magnets are discovered with materials informatics method. Lastly, although some of the suggested two dimensional materials such $CoMo_2$ and $FeCr_2$ as seen in Figure 8.4 have an unstable formation energy, a high magnetic moment is observed. One can consider that such thermodynamically unstable materials can potentially be stabilized by adsorbing gasses or doping. Therefore, the predicted materials in Figure 8.4 give a good suggestion for the experimental synthesis of strong magnets.

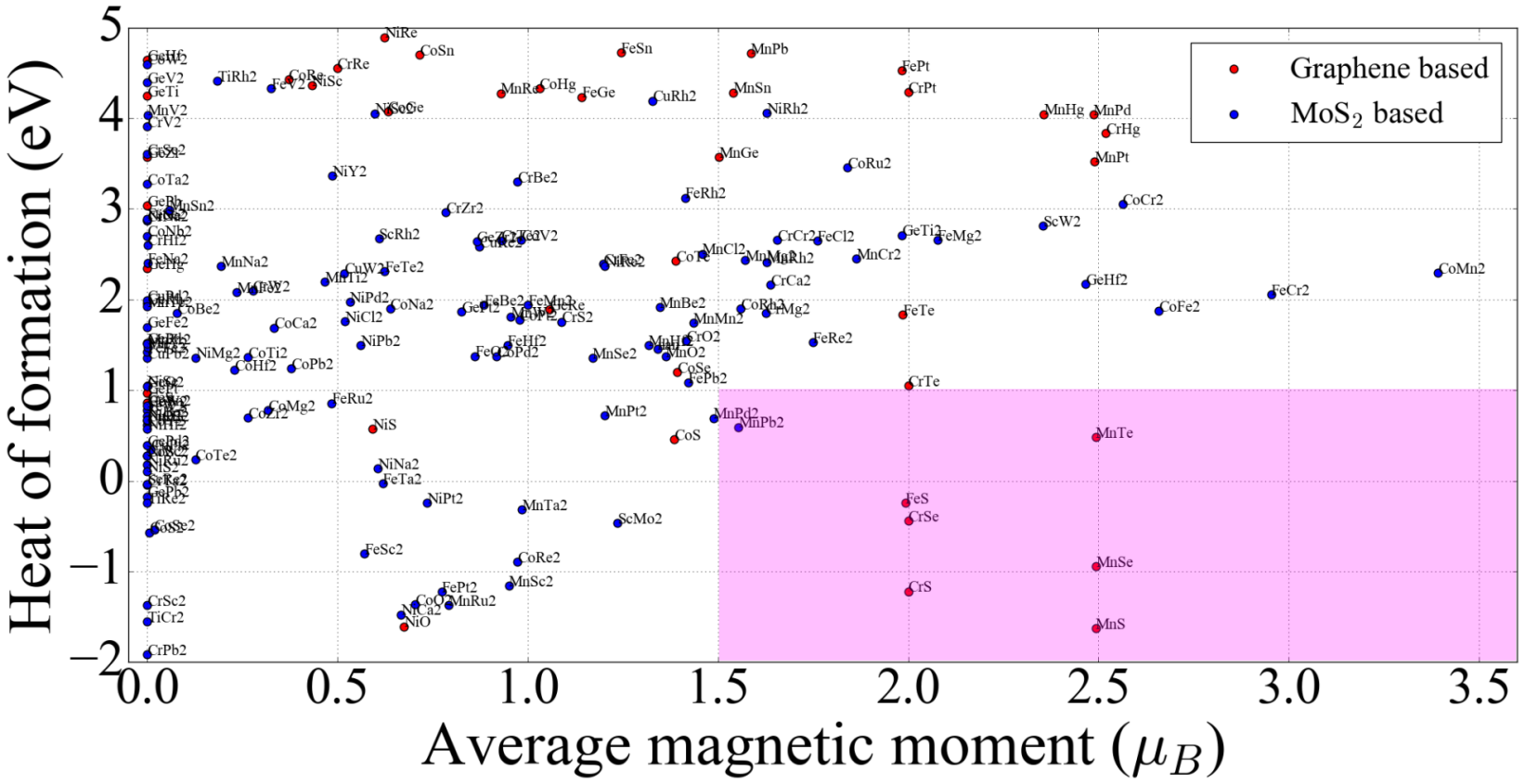

Figure 8.4: Calculated average magnetic moment(μ_B) per atom and formation energy(eV) of the predicted 2D materials listed in Table 8.1 and Table 8.2 The pink-filled area indicates potential two dimensional magnetic materials which have both stable energy state($<$1 eV in formation energy) and high magnetic moment($>1.5\mu_B$).

Table 8.1: Predicted new AB type two dimensional magnets.

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
1	CoCd	5.31	1.00	8.41	4.86
2	CoCr	7.47	1.49	6.60	3.81
3	CoGe	4.07	0.63	7.35	4.24
4	CoHf	7.46	0.44	6.96	4.02
5	CoHg	4.33	1.03	8.45	4.88
6	CoMo	9.11	0.00	6.60	3.81
7	CoO	-2.52	1.41	5.71	3.30
8	CoPb	5.32	0.78	8.41	4.86
9	CoPt	5.28	1.41	7.25	4.19
10	CoRe	4.43	0.37	6.91	3.99
11	CoRh	6.36	1.55	6.98	4.03
12	CoS	0.46	1.38	6.97	4.02
13	CoSc	5.15	0.00	6.77	3.91
14	CoSe	1.20	1.39	7.37	4.25
15	CoSn	4.70	0.72	8.14	4.70
16	CoTa	9.49	0.90	6.84	3.95
17	CoTe	2.43	1.39	8.01	4.63
18	CoTi	6.63	0.48	6.49	3.75
19	CoV	7.22	1.00	6.46	3.73
20	CoW	10.67	0.00	6.67	3.85
21	CoZr	6.00	0.42	7.04	4.07
22	CrCd	5.11	2.54	9.06	5.23
23	CrCr	9.53	5.00	8.42	4.86
24	CrHf	11.34	1.38	7.72	4.46
25	CrHg	3.84	2.52	8.83	5.10
26	CrO	-4.72	0.00	5.66	3.27

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
27	CrPt	4.29	2.00	7.42	4.28
28	CrRe	4.55	0.50	7.03	4.06
29	CrRh	5.78	1.58	7.19	4.15
30	CrS	-1.22	2.00	7.12	4.11
31	CrSe	-0.44	2.00	7.45	4.30
32	CrSn	5.15	1.89	8.33	4.81
33	CrTa	12.51	0.49	7.13	4.11
34	CrTe	1.05	2.00	8.16	4.71
35	CrTi	10.02	1.02	7.11	4.11
36	CrV	9.59	0.49	7.01	4.05
37	CrW	11.77	0.00	7.06	4.08
38	CrZr	9.86	1.02	7.63	4.41
39	FeGe	4.23	1.14	7.52	4.34
40	FeHf	8.55	0.07	7.13	4.11
41	FeNb	8.51	0.00	6.65	3.84
42	FeO	-4.27	0.00	5.84	3.37
43	FePb	5.22	1.27	8.41	4.85
44	FePd	5.21	1.90	7.57	4.37
45	FePt	4.53	1.98	7.32	4.23
46	FeRe	88.51	0.00	6.65	3.84
47	FeS	-0.24	1.99	7.19	4.15
48	FeSc	6.65	0.53	7.13	4.11
49	FeSe	0.37	1.98	7.44	4.29
50	FeSn	4.72	1.24	8.22	4.75
51	FeTa	10.18	0.11	6.74	3.89
52	FeTe	1.83	1.99	8.16	4.71

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
53	FeTi	7.40	0.02	6.65	3.84
54	FeV	7.71	0.50	6.44	3.72
55	FeW	10.85	0.34	6.83	3.94
56	FeZr	7.00	0.03	7.14	4.12
57	GeHf	4.64	0.00	7.80	4.50
58	GeHg	2.35	0.00	8.39	4.84
59	GeNb	6.14	0.50	7.67	4.43
60	GeO	-80.53	0.51	7.77	4.49
61	GePb	3.03	0.00	8.68	5.01
62	GePd	1.52	0.00	7.37	4.25
63	GePt	0.96	0.00	7.37	4.25
64	GeRe	1.89	1.06	7.35	4.24
65	GeRh	1.96	0.00	7.27	4.20
66	GeS	0.86	0.00	8.14	4.70
67	GeTi	4.24	0.00	7.57	4.37
68	GeV	5.26	0.60	7.44	4.29
69	GeW	8.52	1.00	7.50	4.33
70	GeZr	3.57	0.00	7.90	4.56
71	MnCd	5.41	2.34	9.75	5.63
72	MnCr	9.62	4.56	7.95	4.59
73	MnGe	3.57	1.50	7.40	4.27
74	MnHf	10.06	0.56	7.31	4.22
75	MnHg	4.04	2.36	9.46	5.46
76	MnNb	9.22	0.04	7.13	4.11
77	MnO	-4.59	2.45	6.15	3.55
78	MnPb	4.72	1.59	8.63	4.98

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
79	MnPd	4.04	2.49	7.65	4.41
80	MnPt	3.52	2.49	7.56	4.37
81	MnRe	4.27	0.93	6.94	4.01
82	MnRn	93.48	1.48	7.23	4.18
83	MnS	-1.63	2.49	7.28	4.20
84	MnSc	8.03	1.24	7.77	4.49
85	MnSe	-0.95	2.49	7.61	4.39
86	MnSn	4.28	1.54	8.48	4.90
87	MnTe	0.48	2.49	8.24	4.76
88	MnTi	8.63	0.51	7.03	4.06
89	MnV	8.70	0.00	6.35	3.67
90	MnW	11.23	0.42	6.91	3.99
91	MnZr	8.47	0.51	7.57	4.37
92	NiCr	6.88	2.00	6.87	3.97
93	NiGe	2.87	0.00	7.02	4.05
94	NiMo	9.83	1.11	7.17	4.14
95	NiO	-1.61	0.67	5.73	3.31
96	NiPt	5.54	0.82	7.26	4.19
97	NiRe	4.89	0.62	7.02	4.05
98	NiS	0.57	0.59	6.96	4.02
99	NiSc	4.36	0.43	7.03	4.06
100	NiSe	1.05	0.00	7.29	4.21
101	NiTi	6.16	0.99	6.84	3.95
102	NiV	6.94	1.49	6.80	3.92
103	NiW	10.97	1.03	7.06	4.08
104	NiZr	5.79	0.93	7.32	4.23

Table 8.2: Predicted new AB_2 type two dimensional magnets.

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Lattice constant Y
1	$CoBe_2$	1.85	0.08	2.94	2.54
2	$CoCa_2$	1.69	0.33	4.25	3.68
3	$CoCl_2$	-2.12	1.00	3.50	3.03
4	$CoCr_2$	3.05	2.56	3.14	2.72
5	$CoFe_2$	1.87	2.66	3.16	2.74
6	$CoHf_2$	1.23	0.23	3.40	2.94
7	$CoMg_2$	0.78	0.32	3.23	2.80
8	$CoMn_2$	2.30	3.39	3.30	2.86
9	$CoNa_2$	1.90	0.64	3.95	3.42
10	$CoNb_2$	2.70	0.00	3.29	2.85
11	CoO_2	-1.36	0.70	2.68	2.32
12	$CoPb_2$	1.24	0.38	3.87	3.35
13	$CoPd_2$	1.37	0.92	3.30	2.86
14	$CoPt_2$	1.77	0.98	3.30	2.86
15	$CoRe_2$	-0.89	0.97	3.13	2.71
16	$CoRh_2$	1.90	1.56	3.16	2.74
17	$CoRu_2$	3.46	1.84	3.13	2.71
18	CoS_2	-0.57	0.00	3.32	2.87
19	$CoSc_2$	0.27	0.00	3.46	2.99
20	$CoSe_2$	-0.54	0.02	3.46	3.00
21	$CoSn_2$	0.83	0.00	3.75	3.25
22	$CoTa_2$	3.27	0.00	3.26	2.83
23	$CoTe_2$	0.23	0.13	4.01	3.48
24	$CoTi_2$	1.36	0.26	3.23	2.80
25	CoV_2	2.66	0.98	3.16	2.74
26	CoW_2	4.60	0.00	3.16	2.74

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
27	CoY_2	0.61	0.00	3.76	3.26
28	$CoZr_2$	0.70	0.26	3.49	3.02
29	$CrBe_2$	3.30	0.97	3.04	2.63
30	$CrCa_2$	2.16	1.64	4.22	3.66
31	$CrCl_2$	-3.12	1.33	3.51	3.04
32	$CrCr_2$	2.65	1.65	3.23	2.80
33	$CrFe_2$	2.39	1.20	3.13	2.71
34	$CrGe_2$	-29.19	1.99	7.60	4.39
35	$CrHf_2$	2.60	0.00	3.40	2.94
36	$CrKr_2$	3.07	1.91	5.18	4.48
37	$CrMg_2$	1.46	1.34	3.41	2.95
38	$CrNa_2$	1.85	1.63	4.11	3.56
39	$CrNb_2$	2.87	0.00	3.32	2.88
40	CrO_2	-4.88	0.00	2.61	2.26
41	$CrPb_2$	1.55	1.42	3.98	3.45
42	CrS_2	-1.92	0.00	3.20	2.77
43	$CrSc_2$	1.75	1.09	3.49	3.02
44	$CrSe_2$	-1.37	0.00	3.37	2.92
45	$CrTa_2$	3.60	0.00	3.30	2.86
46	$CrTe_2$	-0.05	0.00	3.65	3.16
47	CrV_2	2.65	0.93	3.16	2.74
48	CrW_2	3.91	0.00	3.16	2.74
49	$CrZr_2$	2.09	0.28	3.49	3.02
50	$CuHf_2$	2.96	0.78	3.47	3.00
51	$CuPb_2$	0.66	0.00	3.87	3.36
52	$CuPd_2$	1.35	0.00	3.30	2.86

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
53	$CuPt_2$	1.99	0.00	3.26	2.83
54	$CuRe_2$	0.35	0.01	3.15	2.73
55	$CuRh_2$	2.58	0.87	3.18	2.75
56	$CuRu_2$	4.19	1.33	3.11	2.70
57	$CuTa_2$	5.06	0.76	3.31	2.87
58	CuW_2	6.18	0.00	3.21	2.78
59	$FeBe_2$	2.29	0.52	2.94	2.55
60	$FeCa_2$	1.94	0.88	4.16	3.60
61	$FeCl_2$	-2.79	1.33	3.48	3.02
62	$FeCr_2$	2.65	1.76	3.01	2.60
63	$FeFe_2$	2.06	2.95	3.16	2.74
64	$FeHf_2$	1.42	0.00	3.40	2.94
65	$FeMg_2$	1.49	0.95	3.56	3.08
66	$FeMn_2$	2.66	2.08	3.22	2.79
67	$FeNa_2$	1.94	1.00	3.97	3.44
68	$FeNb_2$	2.40	0.00	3.32	2.88
69	FeO_2	-2.44	0.64	2.65	2.29
70	$FePb_2$	1.38	0.86	3.90	3.38
71	$FePt_2$	1.08	1.42	3.33	2.88
72	$FeRe_2$	-1.23	0.77	3.14	2.72
73	$FeRh_2$	1.53	1.75	3.23	2.80
74	$FeRu_2$	3.12	1.41	3.14	2.72
75	$FeSc_2$	0.86	0.48	3.49	3.02
76	$FeSe_2$	-0.80	0.57	3.50	3.03
77	$FeTa_2$	2.89	0.00	3.25	2.81
78	$FeTe_2$	-0.03	0.62	3.80	3.29

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
79	FeV_2	2.31	0.62	3.16	2.74
80	FeW_2	4.33	0.33	3.17	2.75
81	$FeZr_2$	0.78	0.00	3.46	2.99
82	$GeFe_2$	10.71	1.98	3.26	2.83
83	$GeHf_2$	1.70	0.00	3.39	2.94
84	$GeMn_2$	2.17	2.47	3.29	2.85
85	GeO_2	-2.65	0.00	2.98	2.58
86	$GePb_2$	1.04	0.00	3.87	3.36
87	$GePd_2$	-0.17	0.00	3.20	2.77
88	$GePt_2$	0.39	0.00	3.20	2.77
89	$GeTi_2$	1.87	0.83	3.26	2.83
90	GeV_2	2.71	1.98	3.26	2.83
91	GeW_2	4.40	0.00	3.13	2.71
92	$GeZr_2$	0.83	0.00	3.43	2.97
93	$MnBe_2$	2.64	0.87	2.94	2.54
94	$MnCa_2$	1.91	1.35	4.19	3.63
95	$MnCl_2$	-3.51	1.67	3.53	3.06
96	$MnCr_2$	2.50	1.46	3.11	2.69
97	$MnFe_2$	2.45	1.86	3.07	2.66
98	$MnHf_2$	2.08	0.24	3.39	2.93
99	$MnKr_2$	2.87	1.67	5.58	4.83
100	$MnMg_2$	1.49	1.32	3.59	3.11
101	$MnMn_2$	2.44	1.57	3.20	2.77
102	$MnNa_2$	1.74	1.44	4.02	3.48
103	$MnNb_2$	2.36	0.19	3.26	2.82
104	MnO_2	-3.49	0.32	2.62	2.27

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
105	$MnPb_2$	1.37	1.36	3.90	3.38
106	$MnPd_2$	0.59	1.55	3.33	2.88
107	$MnPt_2$	0.69	1.49	3.26	2.82
108	$MnRh_2$	0.72	1.20	3.15	2.73
109	$MnRu_2$	2.41	1.63	3.15	2.73
110	MnS_2	-1.37	0.79	3.33	2.89
111	$MnSc_2$	70.54	0.90	3.46	3.00
112	$MnSe_2$	-1.16	0.95	3.50	3.03
113	$MnSn_2$	1.35	1.17	3.90	3.38
114	$MnTa_2$	2.99	0.06	3.18	2.75
115	$MnTe_2$	-0.32	0.98	3.80	3.29
116	$MnTi_2$	1.93	0.00	3.20	2.77
117	MnV_2	2.19	0.47	3.06	2.65
118	MnW_2	4.04	0.00	3.18	2.75
119	MnY_2	1.80	0.96	3.84	3.33
120	$MnZr_2$	1.47	0.00	3.43	2.97
121	$NiBe_2$	1.51	0.00	2.82	2.44
122	$NiCa_2$	0.71	0.00	4.06	3.52
123	$NiCl_2$	-1.48	0.67	3.43	2.97
124	$NiHf_2$	1.76	0.52	3.41	2.95
125	$NiMg_2$	0.57	0.00	3.43	2.97
126	$NiNa_2$	1.36	0.13	3.84	3.33
127	NiO_2	0.13	0.61	2.69	2.33
128	$NiPb_2$	0.67	0.00	3.76	3.25
129	$NiPd_2$	1.50	0.56	3.23	2.80
130	$NiPt_2$	1.97	0.53	3.23	2.80

No.	Material	Formation energy (eV)	Average magnetic moment (μB)	Lattice constant X	Y
131	$NiRe_2$	-0.24	0.73	3.15	2.73
132	$NiRh_2$	2.37	1.20	3.15	2.73
133	$NiRu_2$	4.06	1.63	3.15	2.73
134	NiS_2	0.18	0.00	3.27	2.83
135	$NiSc_2$	0.10	0.00	3.43	2.97
136	$NiTa_2$	4.05	0.60	3.28	2.84
137	NiW_2	5.27	0.00	3.18	2.75
138	NiY_2	0.27	0.00	3.77	3.26
139	$ScHf_2$	3.37	0.49	3.70	3.20
140	$ScMo_2$	5.18	1.87	3.47	3.01
141	$ScRe_2$	-0.47	1.24	3.34	2.89
142	$ScRh_2$	-0.03	0.00	3.38	2.92
143	$ScRu_2$	2.67	0.61	3.34	2.89
144	$ScTa_2$	5.24	0.35	3.54	3.06
145	ScW_2	6.08	0.88	3.41	2.95
146	$TiCr_2$	2.81	2.35	3.31	2.87
147	$TiRe_2$	-1.55	0.00	3.25	2.81
148	$TiRh_2$	-0.25	0.00	3.28	2.84
149	$TiTa_2$	4.41	0.18	3.41	2.95
150	TiW_2	5.10	0.00	3.31	2.87

Chapter 9

Conclusions and future research

By the obtained research results, conclusion can be summarized by the following:

- **Boron based two dimensional materials**

  Boron based two dimensional materials are explored and evaluated by using first principle calculations. Hexagonal boron nitride (h-BN) like B-*M* (*M*=P, As, Sb and Bi) materials are designed and revealed to have flat-surface structure with energetically stability, which indicates to be able to synthesize experimentally. Designed novel B-As, B-Sb and B-Bi single layer two dimensional materials have narrow band gap as same as the previously explored B-P single layer two dimensional materials. This narrow band gap property indicates to be a potential candidate for semiconductor material. Those band gaps are revealed to have strong correlation to Allen electronegativity, which can be expanded in Group IV and Group Ⅲ-V binary honeycomb structured compounds. This decline can be useful to design single layer two dimensional semiconductor. B-Sb and B-Bi materials are turned out to have reactive site to hydrogen atom by the evaluation of adsorption energy, which is potential property to consider the use Additionally, bilayer structure is also evaluated in the views of interlayer correlation and electronic structure. All designed materials have layered structure: B-N, B-P and B-As materials are formed interlayer binding caused by physisorption, while interlayer binding of B-Sb and B-Bi materials is originated from chemisorption. Electronic structure differs how layers are designed, indicating that the controlling layer structure

can encourage to tune electronic structure of materials, as well as by controlling the composition. Therefore, those obtained results can be useful to design catalysts and semiconductor within two dimensional materials.

- **Two dimensional magnets**

  Novel two dimensional magnets are explored by machine learning and evaluated by first principle calculations. 216 two dimensional data is collected from open computational database and learned by Gaussian naive Bayes algorithm in order to predict novel two dimensional materials with high magnetic moment. The following 4 descriptors are revealed: atom number of A element$\times$(number of atom in a unit cell), atom number of B element$\times$(number of atom in a unit cell), density of A element$\times$(number of atom in a unit cell), and density of B element$\times$(number of atom in a unit cell). The reversed problem is solved with the same learning algorithm against 746,496 combinations of the 4 descriptors to obtain each corresponding magnetic moment 254 two dimensional magnets are screened by choosing the materials with high magnetic moment within 746,496 materials. These 254 undiscovered two dimensional magnets are evaluated by first principle calculations and 7 stable two dimensional magnets are revealed. As a result, 7 undiscovered two dimensional magnets with high magnetic moment are discovered from 216 original data by using machine learning.

By the comparison of these two research, Materials Informatics approach is more focused on exploring materials as “magnetic moment” is specifically set as target in “Two dimensional magnets” research. On the other hand, “Boron based two dimensional materials” is more conventional as they designed as the reference of previous research in “B-N” and “B-P” two dimensional materials. Therefore, these research provides the superiority of Material Informatics for designing novel functional materials. For the future research, feedback phase should be considered to improve accuracy of material prediction(See blue arrows in Figure 9.1). As mentioned in Chapter 6, dispersed data is key for successful prediction, therefore, adding lacking data(including failure data) to cover whole data dimension in

database is potential for feedback.

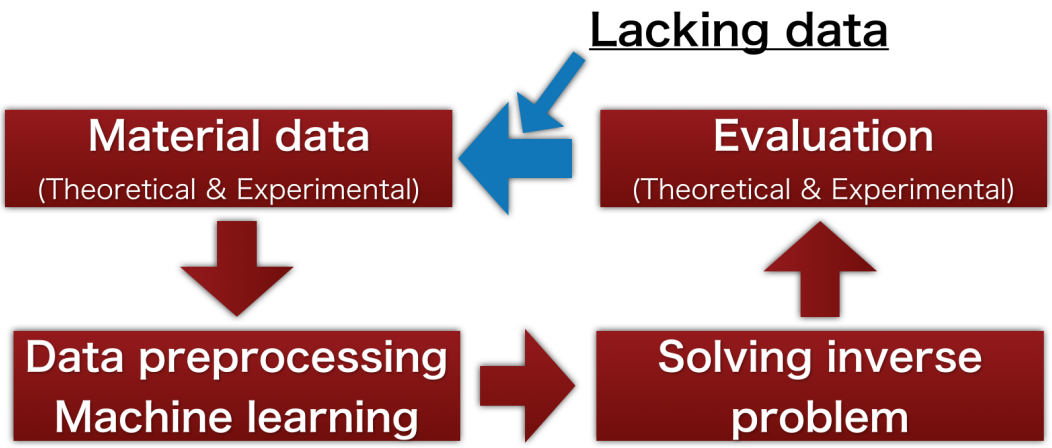

Figure 9.1: The scheme of workflow of Materials Informatics. Blue arrows indicate the potential workflow of future research

The achieved researches still stands inside the frame of theoretical study. The ultimate goal of material science is developing materials in practical use. Simulations, including first principle calculations, can be the potential tool to generate material data because it can provide large quantity of data compared with experiment. However, simulation result can provide the one-sided information that cannot include actual condition of materials. For example, the descriptors of catalysts experimental data of oxidation coupling of methane(OCM) reaction revealed to have the terms of experimental condition[14]. Thus, attempting Materials Informatics approach on experimental data would be the next step of developing novel material in practical use.

Acknowledgements

Firstly, I would like to express my sincere gratitude to Dr. Keisuke Takahashi for mentorship of my Ph D. study and related research. I was not familiar with neither theoretical chemistry nor data science in the beginning of my Ph D. course, however, he kindly invited me entering those new field of research. He support me by giving technical education to improve my research quality, and providing me new sight of research aspects.

Dr. Prof. Kazutoshi Gohara welcomed my transfer from Division of Materials Science and Engineering to Division of Applied Physics due to Dr. Takahashi's transfer. Although I could not be involved with his research deeply because of the time limitation of my doctoral study, he kindly understands my circumstance and encourages me to proceed with my original research. I feel regretful that I could not publish any research achievement with him during my doctoral course, however, I would hopefully like to help his research in the sight of my research field in the near future.

Dr. Prof. Yuzuru Tanaka invited me the world of computer science and provided me a good environment for research. Especially, I would greatly appreciate him to advise me the importance of "commitment" in my life when he told me I was lost for my first career after my graduation. Now I start to go on to researcher's career because of his advice and encouragement.

I really appreciate Dr. Imura Hajime for his comments on my research from the sight of computer science. He also used to be a good advisor of not only my research but also life in Hokkaido and my personal interests.

I also appeciate the members of Laboratory of Advanced Material(LOAM) at

Division of Materials Science and Engineering. Prof. Dr. Somei Ohnuki and Prof. Dr. Naoyuki Hashimoto welcomed me to enter their laboratory for the first year of my doctoral course. It had just been for a year I spent LOAM, however, I somehow learned how research laboratory works. It was a great experience for me.

Last but not the least, I greatly thank to my family; my father, Masayuki Miyazato, my mother, Hitomi Miyazato, and my brother, Kenji Miyazato. They always encourage and support me when I am at the crossroads in my life.

Bibliography

[1] H Şahin, S Cahangirov, M Topsakal, E Bekaroglu, E Akturk, R Tugrul Senger, and Salim Ciraci. Monolayer honeycomb structures of group-IV elements and III-V binary compounds: First-principles calculations. *Physical Review B*, 80(15):155453, Oct 2009.

[2] Filip A. Rasmussen and Kristian S. Thygesen. Computational 2D Materials Database: Electronic Structure of Transition-Metal Dichalcogenides and Oxides. *The Journal of Physical Chemistry C*, 119(23):13169–13183, 2015.

[3] Yifei Mo, Shyue Ping Ong, and Gerbrand Ceder. First principles study of the Li10GeP2S12 lithium super ionic conductor material. *Chemistry of Materials*, 24(1):15–17, 2011.

[4] The NOMAD Laboratory webpage. `https://nomad-coe.eu/`. (Accessed on 10/20/2018).

[5] AiiDA – Features. `http://www.aiida.net/features/`. (Accessed on 11/27/2018).

[6] Xiao-Gang Lu. Remarks on the recent progress of Materials Genome Initiative. *Science Bulletin*, 60(22):1966–1968, Nov 2015.

[7] About MI2I — Materials Research by Information Integration Initiative. `http://www.nims.go.jp/MII-I/en/about/index_m.html`. (Accessed on 09/20/2018).

[8] Motoaki Nishijima, Takuya Ootani, Yuichi Kamimura, Toshitsugu Sueki, Shogo Esaki, Shunsuke Murai, Koji Fujita, Katsuhisa Tanaka, Koji Ohira,

Yukinori Koyama, et al. Accelerated discovery of cathode materials with prolonged cycle life for lithium-ion battery. *Nature communications*, 5:4553, 2014.

[9] Yoyo Hinuma, Taisuke Hatakeyama, Yu Kumagai, Lee A Burton, Hikaru Sato, Yoshinori Muraba, Soshi Iimura, Hidenori Hiramatsu, Isao Tanaka, Hideo Hosono, et al. Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis. *Nature communications*, 7:11962, 2016.

[10] Prasanna V Balachandran, Benjamin Kowalski, Alp Sehirlioglu, and Turab Lookman. Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nature communications*, 9, 2018.

[11] BASF and Citrine Informatics collaborate to use artificial intelligence to develop new catalyst technology. `https://www.basf.com/us/en/company/news-and-media/news-releases/2018/06/P-US-18-075.html`. (Accessed on 11/27/2018).

[12] Toyota Research Institute Brings Artificial Intelligence to the Hunt for New Materials — Toyota USA Newsroom. `https://pressroom.toyota.com/releases/tri+artificial+intelligence+new+materials+march30.htm`. (Accessed on 11/27/2018).

[13] Newswitch:the frontier of Materials Informatics (in Japanese). `https://newswitch.jp/p/11561`. (Accessed on 10/20/2018).

[14] Keisuke Takahashi, Itsuki Miyazato, Shun Nishimura, and Junya Ohyama. Unveiling Hidden Catalysts for the Oxidative Coupling of Methane based on Combining Machine Learning with Literature Data. *ChemCatChem*, 10(15):3223–3228, 2018.

[15] Gene Ontology Consortium — Gene Ontology Consortium. `http://www.geneontology.org/`. (Accessed on 01/01/2019).

[16] Lauren Takahashi, Itsuki Miyazato, and Keisuke Takahashi. Redesigning the Materials and Catalysts Database Construction Process Using Ontologies. *Journal of chemical information and modeling*, 58(9):1742–1754, 2018.

[17] Janis Timoshenko, Deyu Lu, Yuewei Lin, and Anatoly I Frenkel. Supervised machine-learning-based determination of three-dimensional structure of metallic nanoparticles. *The journal of physical chemistry letters*, 8(20):5091–5098, 2017.

[18] Chen Zheng, Kiran Mathew, Chi Chen, Yiming Chen, Hanmei Tang, Alan Dozier, Joshua J Kas, Fernando D Vila, John J Rehr, Louis FJ Piper, et al. Automated generation and ensemble-learned matching of X-ray absorption spectra. *npj Computational Materials*, 4(1):12, 2018.

[19] Kiran Mathew, Chen Zheng, Donald Winston, Chi Chen, Alan Dozier, John J Rehr, Shyue Ping Ong, and Kristin A Persson. High-throughput computational X-ray absorption spectroscopy. *Scientific data*, 5:180151, 2018.

[20] Janis Timoshenko, Avik Halder, Bing Yang, Soenke Seifert, Michael J Pellin, Stefan Vajda, and Anatoly I Frenkel. Subnanometer Substructures in Nanoassemblies Formed from Clusters under a Reactive Atmosphere Revealed Using Machine Learning. *The Journal of Physical Chemistry C*, 122(37):21686–21693, 2018.

[21] José Francisco Díez-Pastor, Susana Esther Jorge-Villar, Álvar Arnaiz-González, César Ignacio García-Osorio, Yael Díaz-Acha, Marc Campeny, Josep Bosch, and Joan Carles Melgarejo. Machine learning algorithms applied to R aman spectra for the identification of variscite originating from the mining complex of G avà. *Journal of Raman Spectroscopy*, 2018.

[22] Tharanga K Wijethunga, Jelena Stojaković, Michael A Bellucci, Xingyu Chen, Allan S Myerson, and Bernhardt L Trout. General Method for

the Identification of Crystal Faces Using Raman Spectroscopy Combined with Machine Learning and Application to the Epitaxial Growth of Acetaminophen. *Langmuir*, 34(33):9836–9846, 2018.

[23] Kiyotaka Asakura, Hitoshi Abe, and Masao Kimura. The challenge of constructing an international XAFS database. *Journal of Synchrotron Radiation*, 25(4):967–971, Jul 2018.

[24] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y. Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov. Electric Field Effect in Atomically Thin Carbon Films. *Science*, 306(5696):666–669, 2004.

[25] Andre K Geim and Konstantin S Novoselov. *The rise of graphene*, pages 11–19. World Scientific, 2010.

[26] Ferdinand Hof, Alessandro Boni, Giovanni Valenti, Kai Huang, Francesco Paolucci, and Alain Pénicaud. From Food Waste to Efficient Bifunctional Nonprecious Electrocatalyst. *Chemistry – A European Journal*, 23(61):15283–15288.

[27] Michael Ashton, Joshua Paul, Susan B. Sinnott, and Richard G. Hennig. Topology-Scaling Identification of Layered Solids and Stable Exfoliated 2D Materials. *Phys. Rev. Lett.*, 118:106101, Mar 2017.

[28] The Nobel Prize in Physics 2010. `https://www.nobelprize.org/prizes/physics/2010/summary/`. (Accessed on 09/20/2018).

[29] Hans Peter Boehm, Ralph Setton, and Eberhard Stumpp. Nomenclature and terminology of graphite intercalation compounds (IUPAC Recommendations 1994). *Pure and Applied Chemistry*, 66(9):1893–1901, 1994.

[30] H. P. Boehm, A. Clauss, G. O. Fischer, and U. Hofmann. Das Adsorptionsverhalten sehr dünner Kohlenstoff-Folien. *Zeitschrift für anorganische und allgemeine Chemie*, 316(34):119–127.

[31] A. K. Geim. Graphene: Status and Prospects. *Science*, 324(5934):1530–1534, 2009.

[32] Daniel R Cooper, Benjamin D'Anjou, Nageswara Ghattamaneni, Benjamin Harack, Michael Hilke, Alexandre Horth, Norberto Majlis, Mathieu Massicotte, Leron Vandsburger, Eric Whiteway, et al. Experimental review of graphene. *ISRN Condensed Matter Physics*, 2012, 2012.

[33] O. A. Shenderova, V. V. Zhirnov, and D. W. Brenner. Carbon Nanostructures. *Critical Reviews in Solid State and Materials Sciences*, 27(3-4):227–356, 2002.

[34] Annalisa Fasolino, JH Los, and Mikhail I Katsnelson. Intrinsic ripples in graphene. *Nature materials*, 6(11):858, 2007.

[35] Johan M Carlsson. Graphene: buckle or break. *Nature Materials*, 6(11):801, 2007.

[36] CNR Rao and Urmimala Maitra. Inorganic graphene analogs. *Annual Review of Materials Research*, 45:29–62, 2015.

[37] CNR Rao and K Gopalakrishnan. Borocarbonitrides, B x C y N z: Synthesis, Characterization, and Properties with Potential Applications. *ACS applied materials & interfaces*, 9(23):19478–19494, 2016.

[38] Nitesh kumar, Kota Moses, K. Pramoda, Sharmila N. Shirodkar, Abhishek Kumar Mishra, Umesh V. Waghmare, A. Sundaresan, and C. N. R. Rao. Borocarbonitrides, BxCyNz. *J. Mater. Chem. A*, 1:5806–5821, 2013.

[39] Yi-Hsien Lee, Xin-Quan Zhang, Wenjing Zhang, Mu-Tung Chang, Cheng-Te Lin, Kai-Di Chang, Ya-Chu Yu, Jacob Tse-Wei Wang, Chia-Seng Chang, Lain-Jong Li, et al. Synthesis of large-area MoS2 atomic layers with chemical vapor deposition. *Advanced Materials*, 24(17):2320–2325, 2012.

[40] Yi Zhang, Tay-Rong Chang, Bo Zhou, Yong-Tao Cui, Hao Yan, Zhongkai Liu, Felix Schmitt, James Lee, Rob Moore, Yulin Chen, et al. Direct observation of the transition from indirect to direct bandgap in atomically thin epitaxial MoSe2. *Nature nanotechnology*, 9(2):111–115, 2014.

[41] D Puotinen and RE Newnham. The crystal structure of MoTe2. *Acta Crystallographica*, 14(6):691–692, 1961.

[42] BRUCE E Brown. The crystal structures of WTe2 and high-temperature MoTe2. *Acta Crystallographica*, 20(2):268–274, 1966.

[43] R Tenne, L Margulis, M ea Genut, and G Hodes. Polyhedral and cylindrical structures of tungsten disulphide. *Nature*, 360(6403):444, 1992.

[44] J Pouzet, JC Bernede, A Khellil, H Essaidi, and S Benhida. Preparation and characterization of tungsten diselenide thin films. *Thin Solid Films*, 208(2):252–259, 1992.

[45] Andrea Splendiani, Liang Sun, Yuanbo Zhang, Tianshu Li, Jonghwan Kim, Chi-Yung Chim, Giulia Galli, and Feng Wang. Emerging photoluminescence in monolayer MoS2. *Nano letters*, 10(4):1271–1275, 2010.

[46] Branimir Radisavljevic, Aleksandra Radenovic, Jacopo Brivio, i V Giacometti, and A Kis. Single-layer MoS2 transistors. *Nature nanotechnology*, 6(3):147–150, 2011.

[47] RS Sundaram, M Engel, A Lombardo, R Krupke, AC Ferrari, Ph Avouris, and M Steiner. Electroluminescence in single layer MoS2. *Nano letters*, 13(4):1416–1421, 2013.

[48] Oriol Lopez-Sanchez, Dominik Lembke, Metin Kayci, Aleksandra Radenovic, and Andras Kis. Ultrasensitive photodetectors based on monolayer MoS2. *Nature nanotechnology*, 8(7):497–501, 2013.

[49] Joelson C Garcia, Denille B de Lima, Lucy VC Assali, and Joao F Justo. Group IV graphene-and graphane-like nanosheets. *The Journal of Physical Chemistry C*, 115(27):13242–13246, 2011.

[50] Elisabeth Bianco, Sheneve Butler, Shishi Jiang, Oscar D Restrepo, Wolfgang Windl, and Joshua E Goldberger. Stability and exfoliation of germanane: a germanium graphane analogue. *Acs Nano*, 7(5):4414–4421, 2013.

[51] ME Dávila, Lede Xian, Seymur Cahangirov, Angel Rubio, and Guy Le Lay. Germanene: a novel two-dimensional germanium allotrope akin to graphene and silicene. *New Journal of Physics*, 16(9):095002, 2014.

[52] Shiro Kaneko, Hideaki Tsuchiya, Yoshinari Kamakura, Nobuya Mori, and Matsuto Ogawa. Theoretical performance estimation of silicene, germanene, and graphene nanoribbon field-effect transistors under ballistic transport. *Applied Physics Express*, 7(3):035102, 2014.

[53] F. Reis, G. Li, L. Dudy, M. Bauernfeind, S. Glass, W. Hanke, R. Thomale, J. Schäfer, and R. Claessen. Bismuthene on a SiC substrate: A candidate for a high-temperature quantum spin Hall material. *Science*, 357(6348):287–290, 2017.

[54] Antonis N Andriotis, Ernst Richter, and Madhu Menon. Prediction of a new graphenelike Si 2 BN solid. *Physical Review B*, 93(8):081413, 2016.

[55] L. Debbichi, H. Kim, T. Björkman, O. Eriksson, and S. Lebègue. First-principles investigation of two-dimensional trichalcogenide and sesquichalcogenide monolayers. *Phys. Rev. B*, 93:245307, Jun 2016.

[56] Jun Dai, Ming Li, and Xiao Cheng Zeng. Group IVB transition metal trichalcogenides: a new class of 2D layered materials beyond graphene. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 6(2):211–222, 2016.

[57] Qiyi Zhao, Yaohui Guo, Yixuan Zhou, Zehan Yao, Zhaoyu Ren, Jintao Bai, and Xinlong Xu. Band alignments and heterostructures of monolayer transition metal trichalcogenides MX 3 (M= Zr, Hf; X= S, Se) and dichalcogenides MX 2 (M= Tc, Re; X= S, Se) for solar applications. *Nanoscale*, 10(7):3547–3555, 2018.

[58] Yalong Jiao, Liujiang Zhou, Fengxian Ma, Guoping Gao, Liangzhi Kou, John Bell, Stefano Sanvito, and Aijun Du. Predicting single-layer technetium dichalcogenides (TcX2, X= S, Se) with promising applications in photovoltaics and photocatalysis. *ACS applied materials & interfaces*, 8(8):5385–5392, 2016.

[59] Yingdi Jin, Xingxing Li, and Jinlong Yang. Single layer of MX 3 (M= Ti, Zr; X= S, Se, Te): a new platform for nano-electronics and optics. *Physical Chemistry Chemical Physics*, 17(28):18665–18669, 2015.

[60] Carmen C Mayorga-Martinez, Zdeněk Sofer, Jan Luxa, Štěpán Huber, David Sedmidubsky, Petr Brázda, Lukáš Palatinus, Martin Mikulics, Petr Lazar, Rostislav Medlín, et al. TaS3 Nanofibers: Layered Trichalcogenide for High-Performance Electronic and Sensing Devices. *ACS nano*, 12(1):464–473, 2017.

[61] Shengxue Yang, Minghui Wu, Wanfu Shen, Li Huang, Sefaattin Tongay, Kedi Wu, Bin Wei, Ying Qin, Zhongchang Wang, Chengbao Jiang, et al. Highly Sensitive Polarization Photodetection Using Pseudo-one-dimensional Nb (1-x) TixS3 Alloy. *ACS Applied Materials & Interfaces*, 2018.

[62] Yasir Saeed, Ali Kachmar, and Marcelo A Carignano. First-principles study of the transport properties in bulk and monolayer MX3 (M= Ti, Zr, Hf and X= S, Se) compounds. *The Journal of Physical Chemistry C*, 121(3):1399–1403, 2017.

[63] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.*, 140:A1133–A1138, Nov 1965.

[64] Axel D. Becke. Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics*, 140(18):18A301, 2014.

[65] David C. Langreth and M. J. Mehl. Beyond the local-density approximation in calculations of ground-state electronic properties. *Phys. Rev. B*, 28:1809–1834, Aug 1983.

[66] A. D. Becke. Density-functional exchange-energy approximation with correct asymptotic behavior. *Phys. Rev. A*, 38:3098–3100, Sep 1988.

[67] John P. Perdew, J. A. Chevary, S. H. Vosko, Koblar A. Jackson, Mark R. Pederson, D. J. Singh, and Carlos Fiolhais. Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B*, 46:6671–6687, Sep 1992.

[68] Axel D. Becke. Densityfunctional thermochemistry. III. The role of exact exchange. *The Journal of Chemical Physics*, 98(7):5648–5652, 1993.

[69] John P Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical review letters*, 77(18):3865, 1996.

[70] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The SIESTA method for ab initio order-N materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.

[71] Graeme Henkelman, Andri Arnaldsson, and Hannes Jónsson. A fast and robust algorithm for Bader decomposition of charge density. *Computational Materials Science*, 36(3):354–360, 2006.

[72] Edward Sanville, Steven D Kenny, Roger Smith, and Graeme Henkelman. Improved grid-based algorithm for Bader charge allocation. *Journal of computational chemistry*, 28(5):899–908, 2007.

[73] J e Enkovaara, Carsten Rostgaard, Jens Jørgen Mortensen, Jingzhe Chen, M Dułak, Lara Ferrighi, Jeppe Gavnholt, Christian Glinsvad, V Haikola,

HA Hansen, et al. Electronic structure calculations with GPAW: a real-space implementation of the projectoraugmented-wave method. *Journal of Physics: Condensed Matter*, 22(25):253202, 2010.

[74] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996.

[75] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational materials science*, 6(1):15–50, 1996.

[76] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, Jr. J. A. Montgomery, J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. Gaussian16 Revision B.01, 2016. Gaussian Inc. Wallingford CT.

[77] Paolo Giannozzi, Stefano Baroni, Nicola Bonini, Matteo Calandra, Roberto Car, Carlo Cavazzoni, Davide Ceresoli, Guido L Chiarotti, Matteo Cococcioni, Ismaila Dabo, et al. QUANTUM ESPRESSO: a modular and open-source software project for quantum simulations of materials. *Journal of physics: Condensed matter*, 21(39):395502, 2009.

[78] P Giannozzi, Oliviero Andreussi, T Brumme, O Bunau, M Buongiorno Nardelli, M Calandra, R Car, C Cavazzoni, D Ceresoli, M Cococcioni, et al. Advanced capabilities for materials modelling with Quantum ESPRESSO. *Journal of Physics: Condensed Matter*, 29(46):465901, 2017.

[79] P. E. Blöchl. Projector augmented-wave method. *Phys. Rev. B*, 50:17953–17979, Dec 1994.

[80] Peter E. Blöchl, Clemens J. Först, and Johannes Schimpl. Projector augmented wave method:ab initio molecular dynamics with full wave functions. *Bulletin of Materials Science*, 26(1):33–41, Jan 2003.

[81] J. J. Mortensen, L. B. Hansen, and K. W. Jacobsen. Real-space grid implementation of the projector augmented wave method. *Phys. Rev. B*, 71:035109, Jan 2005.

[82] J Enkovaara, C Rostgaard, J J Mortensen, J Chen, M Dułak, L Ferrighi, J Gavnholt, C Glinsvad, V Haikola, H A Hansen, H H Kristoffersen, M Kuisma, A H Larsen, L Lehtovaara, M Ljungberg, O Lopez-Acevedo, P G Moses, J Ojanen, T Olsen, V Petzold, N A Romero, J Stausholm-Møller, M Strange, G A Tritsaris, M Vanin, M Walter, B Hammer, H Häkkinen, G K H Madsen, R M Nieminen, J K Nørskov, M Puska, T T Rantala, J Schiøtz, K S Thygesen, and K W Jacobsen. Electronic structure calculations with GPAW: a real-space implementation of the projector augmented-wave method. *Journal of Physics: Condensed Matter*, 22(25):253202, 2010.

[83] Code: Bader Charge Analysis. `http://theory.cm.utexas.edu/henkelman/code/bader/`. (Accessed on 10/20/2018).

[84] S. R. Bahn and K. W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.*, 4(3):56–66, MAY-JUN 2002.

[85] Ask Hjorth Larsen, Jens Jørgen Mortensen, Jakob Blomqvist, Ivano E Castelli, Rune Christensen, Marcin Dułak, Jesper Friis, Michael N Groves,

Bjørk Hammer, Cory Hargus, Eric D Hermes, Paul C Jennings, Peter Bjerre Jensen, James Kermode, John R Kitchin, Esben Leonhard Kolsbjerg, Joseph Kubal, Kristen Kaasbjerg, Steen Lysgaard, Jón Bergmann Maronsson, Tristan Maxson, Thomas Olsen, Lars Pastewka, Andrew Peterson, Carsten Rostgaard, Jakob Schiøtz, Ole Schütt, Mikkel Strange, Kristian S Thygesen, Tejs Vegge, Lasse Vilhelmsen, Michael Walter, Zhenhua Zeng, and Karsten W Jacobsen. The atomic simulation environment—a Python library for working with atoms. *Journal of Physics: Condensed Matter*, 29(27):273002, 2017.

[86] Open MPI: Open Source High Performance Computing. `https://www.open-mpi.org/`. (Accessed on 12/31/2018).

[87] FFTW Home Page. `http://www.fftw.org/`. (Accessed on 12/29/2018).

[88] ScaLAPACK — Scalable Linear Algebra PACKage. `http://www.netlib.org/scalapack/`. (Accessed on 12/31/2018).

[89] OpenSSH. `https://www.openssh.com/`. (Accessed on 01/02/2019).

[90] Tutorials — GPAW. `https://wiki.fysik.dtu.dk/gpaw/tutorials/tutorials.html`. (Accessed on 01/06/2019).

[91] Itsuki Miyazato and Keisuke Takahashi. Electronic structure of boron based single and multi-layer two dimensional materials. *Journal of Applied Physics*, 122(10):104302, 2017.

[92] Deniz Çakr, Deniz Kecik, Hasan Sahin, Engin Durgun, and Francois M. Peeters. Realization of a p–n junction in a single layer boron-phosphide. *Phys. Chem. Chem. Phys.*, 17:13013–13020, 2015.

[93] José M Soler, Emilio Artacho, Julian D Gale, Alberto García, Javier Junquera, Pablo Ordejón, and Daniel Sánchez-Portal. The SIESTA method for ab initio order- N materials simulation. *Journal of Physics: Condensed Matter*, 14(11):2745, 2002.

[94] Hendrik J Monkhorst and James D Pack. Special points for Brillouin-zone integrations. *Physical review B*, 13(12):5188, 1976.

[95] L Boldrin, F Scarpa, R Chowdhury, and S Adhikari. Effective mechanical properties of hexagonal boron nitride nanosheets. *Nanotechnology*, 22(50):505702, 2011.

[96] Jin Yu and Wanlin Guo. Strain tunable electronic and magnetic properties of pristine and semihydrogenated hexagonal boron phosphide. *Applied Physics Letters*, 106(4):043107, 2015.

[97] Chuanhong Jin, Fang Lin, Kazu Suenaga, and Sumio Iijima. Fabrication of a Freestanding Boron Nitride Single Layer and Its Defect Assignments. *Phys. Rev. Lett.*, 102:195505, May 2009.

[98] Pekka Pyykkö and Michiko Atsumi. Molecular Single-Bond Covalent Radii for Elements 1–118. *Chemistry–A European Journal*, 15(1):186–197, 2009.

[99] Pekka Pyykkö and Michiko Atsumi. Molecular Double-Bond Covalent Radii for Elements Li–E112. *Chemistry–A European Journal*, 15(46):12770–12779, 2009.

[100] Pekka Pyykkö, Sebastian Riedel, and Michael Patzschke. Triple-Bond Covalent Radii. *Chemistry–A European Journal*, 11(12):3511–3520, 2005.

[101] Tom Leadbetter Cottrell. *The strengths of chemical bonds*. Academic Press, 1958.

[102] Leland C Allen. Electronegativity is the average one-electron energy of the valence-shell electrons in ground-state free atoms. *Journal of the American Chemical Society*, 111(25):9003–9014, 1989.

[103] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.*, 92:246401, Jun 2004.

[104] H Huepen, G Will, F Elf, and VL Solozhenko. Isothermal compression of rhombohedral boron-nitride up to 14 GPa. *Solid State Commun*, 96:1–3, 1995.

[105] A. L. Samuel. Some Studies in Machine Learning Using the Game of Checkers. *IBM Journal of Research and Development*, 3(3):210–229, July 1959.

[106] Alan Stuart and J Keith Ord. *Kendall's Advanced Theory of Statistics, Volume 1: Distribution Theory, Edward Arnold*. 1994.

[107] M Narasimha Murty and V Susheela Devi. *Pattern recognition: An algorithmic approach*. Springer Science & Business Media, 2011.

[108] Welcome — Theano 1.0.0 documentation. `http://deeplearning.net/software/theano/`. (Accessed on 12/31/2018).

[109] TensorFlow. `https://www.tensorflow.org/?hl=ja`. (Accessed on 12/31/2018).

[110] Caffe — Deep Learning Framework. `http://caffe.berkeleyvision.org/`. (Accessed on 12/31/2018).

[111] Preferred Networks. `https://www.preferred-networks.jp/en/`. (Accessed on 01/06/2019).

[112] Chainer: A flexible framework for neural networks. `https://chainer.org/`. (Accessed on 12/31/2018).

[113] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[114] NumPy — NumPy. `http://www.numpy.org/`. (Accessed on 12/31/2018).

[115] SciPy.org — SciPy.org. https://www.scipy.org/. (Accessed on 12/31/2018).

[116] Python Data Analysis Library — pandas: Python Data Analysis Library. https://pandas.pydata.org/. (Accessed on 12/31/2018).

[117] Matplotlib: Python plotting — Matplotlib 3.0.2 documentation. https://matplotlib.org/. (Accessed on 12/31/2018).

[118] Who is using scikit-learn? — scikit-learn 0.20.2 documentation. https://scikit-learn.org/stable/testimonials/testimonials.html. (Accessed on 12/31/2018).

[119] scikit-learn: machine learning in Python — scikit-learn 0.20.2 documentation. https://scikit-learn.org/stable/. (Accessed on 01/04/2019).

[120] Choosing the right estimator — scikit-learn 0.20.2 documentation. https://scikit-learn.org/stable/tutorial/machine_learning_map/index.html. (Accessed on 01/04/2019).

[121] Home - Anaconda. https://www.anaconda.com/. (Accessed on 12/31/2018).

[122] Itsuki Miyazato, Yuzuru Tanaka, and Keisuke Takahashi. Accelerating the discovery of hidden two-dimensional magnets using machine learning and first principle calculations. *Journal of Physics: Condensed Matter*, 30(6):06LT01, jan 2018.

[123] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin A. Persson. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.

[124] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

[125] Keisuke Takahashi and Yuzuru Tanaka. Materials informatics: a journey towards material design and synthesis. *Dalton Trans.*, 45:10497–10499, 2016.

[126] Anubhav Jain, Yongwoo Shin, and Kristin A Persson. Computational predictions of energy materials using density functional theory. *Nature Reviews Materials*, 1(1):15004, 2016.

[127] Angel Rubio, Jennifer L. Corkill, and Marvin L. Cohen. Theory of graphitic boron nitride nanotubes. *Phys. Rev. B*, 49:5081–5084, Feb 1994.

[128] Artem R Oganov and Shigeaki Ono. Theoretical and experimental evidence for a post-perovskite phase of MgSiO 3 in Earth's D layer. *Nature*, 430(6998):445, 2004.

[129] Lei Cheng, Rajeev S. Assary, Xiaohui Qu, Anubhav Jain, Shyue Ping Ong, Nav Nidhi Rajput, Kristin Persson, and Larry A. Curtiss. Accelerating Electrolyte Discovery for Energy Storage with High-Throughput Screening. *The Journal of Physical Chemistry Letters*, 6(2):283–291, 2015. PMID: 26263464.

[130] Stefano Curtarolo, Wahyu Setyawan, Shidong Wang, Junkai Xue, Kesong Yang, Richard H Taylor, Lance J Nelson, Gus LW Hart, Stefano Sanvito, Marco Buongiorno-Nardelli, et al. AFLOWLIB. ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science*, 58:227–235, 2012.

[131] David D Landis, Jens S Hummelshoj, Svetlozar Nestorov, Jeff Greeley, Marcin Dulak, Thomas Bligaard, Jens K Norskov, and Karsten W Jacobsen.

The computational materials repository. *Computing in Science & Engineering*, 14(6):51–57, 2012.

[132] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, and Kristin a. Persson. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.

[133] D. Kerkmann, D. Pescia, and R. Allenspach. Two-dimensional magnet at Curie temperature: Epitaxial layers of Co on Cu(100). *Phys. Rev. Lett.*, 68:686–689, Feb 1992.

[134] Bevin Huang, Genevieve Clark, Efrén Navarro-Moratalla, Dahlia R Klein, Ran Cheng, Kyle L Seyler, Ding Zhong, Emma Schmidgall, Michael A McGuire, David H Cobden, et al. Layer-dependent ferromagnetism in a van der Waals crystal down to the monolayer limit. *Nature*, 546(7657):270, 2017.

[135] Wei-Bing Zhang, Qian Qu, Peng Zhu, and Chi-Hang Lam. Robust intrinsic ferromagnetism and half semiconductivity in stable two-dimensional single-layer chromium trihalides. *Journal of Materials Chemistry C*, 3(48):12457–12468, 2015.

[136] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth, and B. I. Lundqvist. Van der Waals Density Functional for General Geometries. *Phys. Rev. Lett.*, 92:246401, Jun 2004.